OPTIMIZATION APPROACHES FOR PLANNING EXTERNAL BEAM RADIOTHERAPY

A Thesis
Presented to
The Academic Faculty

by

Halil Ozan Gözbaşı

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2010

OPTIMIZATION APPROACHES FOR PLANNING EXTERNAL BEAM RADIOTHERAPY

Approved by:

Shabbir Ahmed, Co-advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Martin W. P. Savelsbergh, Co-advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Pınar Keskinocak
School of Industrial and Systems Engineering
Georgia Institute of Technology

Eduard Schreibmann
Department of Radiation Oncology
Emory School of Medicine

Joel Sokol
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 19 May 2010

To Ümran Şefik,

my primary school teacher

ACKNOWLEDGEMENTS

I would like to begin by thanking to my advisors Dr. Shabbir Ahmed and Dr. Martin Savelsbergh, who provided me the research opportunity and guidance during my Ph.D. program. They helped me to develop my skills both as a researcher and a future practitioner of operations research.

I would like to thank Dr. Ian Crocker, Dr. Tim Fox and Dr. Eduard Schreibmann for sharing their knowledge and clinical experience in radiation treatment. I am grateful to Georgia Institute of Technology and Emory University School of Medicine for their financial support.

I have had the pleasure to meet Dr. Pınar Keskinocak and Dr. Joel Sokol, and would like to thank for their time and effort in serving on my proposal and defense committees.

I would like to express my deepest gratitude to my mom and dad, Hatime and Nazmi Gözbaşı, for their patience and letting me study far away from home since childhood. I am out of words to describe how valuable it is to hear their warm voice and encouragement on the phone during hard times. I would like to thank my brother Onur for being a role model for me, and my sister Berna for her generous love.

I feel lucky to be roommates with Gökhan Samur and Emrah Uyar. You have been my brothers here. I have had a great time in Atlanta by building invaluable friendships. Aykağan Ak, Kaya Demir, Serhan Duran, Ali Ekici, Adem Esen, Gökçe Gürün, Cihan İkikat, Burak Karacık, Gonca Karakuş, Orhan Karzan, Mazlum Koşma, Örsan Özener, Evren Özkaya, Şahin Tetik, Burç Tosyalı and Erdem Uçar are a few to name here.

Finally, the biggest thanks go to my fiancee Tuba. You went through all the stressful times with me in my final year. It would be much harder to finish my thesis without your love, patience and support. I am extremely lucky to meet you here and you are the best present Georgia Tech gave me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Cancer begins when cells grow out of control as a result of damage to their DNA. These abnormal cells can invade healthy tissue and form tumors in various parts of the body. Chemotherapy, immunotherapy, surgery and radiotherapy are the most common treatment methods for cancer. According to American Cancer Society about half of the cancer patients receive a form of radiation therapy at some stage. External beam radiotherapy is delivered from outside the body and aimed at cancer cells to damage their DNA making them unable to divide and reproduce. The beams travel through the body and may damage nearby healthy tissue unless carefully planned. Therefore, the goal of treatment plan optimization is to find the best system parameters to deliver sufficient dose to target structures while avoiding damage to healthy tissue. This thesis investigates optimization approaches for two external beam radiation therapy techniques: Intensity-Modulated Radiation Therapy (IMRT) and Volumetric-Modulated Arc Therapy (VMAT).

We develop automated treatment planning technology for IMRT that produces several high-quality treatment plans satisfying provided clinical requirements in a single invocation and without human guidance. A novel bi-criteria scoring based beam selection algorithm is part of the planning system and produces better plans compared to those produced using a well-known scoring-based algorithm. Our algorithm is very efficient and finds the beam configuration at least ten times faster than an exact integer programming approach. Solution times range from 2 minutes to 15 minutes which is clinically acceptable.

With certain cancers, especially lung cancer, a patient's anatomy changes during treatment. These anatomical changes need to be considered in treatment planning. Fortunately, recent advances in imaging technology can provide multiple images of the treatment region taken at different points of the breathing cycle, and deformable image registration algorithms can accurately link these images making it possible to track an individual voxel

during the entire breathing cycle. This allows the development of optimization models that generate treatment plans that deliver radiation in multiple phases. Our model finds optimal fluence maps for each phase of the breathing cycle simultaneously by considering the overall dose delivered to patient. Because the optimization exploits the specific opportunities provided in each of the phases, better treatment plans are obtained. A computational study of a real lung case shows that the tumor coverage can be improved from only 51% using single-phase gating to 96% using five-phase gating.

VMAT is a recent radiation treatment technology which is based on continuous rotation of the radiation source around the patient. VMAT has the potential to provide treatments in less time compared to other delivery techniques which enhances patient comfort and allows for the treatment of more patients. A treatment planning system has to decide how the beam shapes and dose rates change during the rotation of the radiation source. We develop two treatment planning approaches for VMAT. The first approach finds shapes and dose rates separately in a two-stage algorithm. Although this approach produces treatment plans extremely fast their quality is not clinically acceptable. The second approach is based on a large-scale mixed-integer programming model that optimizes shapes and dose rates simultaneously. As solving the model directly is computationally prohibitive, we develop a heuristic approach which solves the model multiple times on a reduced set of decision variables. We derive valid inequalities that not only decrease the solution times but also allows us to get better integer solutions within specified time limits. Computational studies on a spinal tumor and a prostate tumor case produce clinically acceptable results.

CHAPTER I

INTRODUCTION

Cancer begins when cells grow out of control as a result of damage to their DNA. These abnormal cells can invade healthy tissue and form tumors in various parts of the body. About 1.5 million people are diagnosed with cancer each year and about 11 million people live with cancer at any point in time. Chemotherapy, immunotherapy, surgery and radiotherapy are the most common treatment methods for curing cancer. According to American Cancer Society about half of the cancer patients receive a form of radiation therapy at some stage. There are two ways to deliver radiation. Internal radiation is delivered from inside the body by surgically placing radioactive material directly in the tumor, while external beam radiation is delivered from outside the body by using a machine. The external beams are aimed to cancer cells to damage their DNA and make them unable to divide and reproduce. These beams travel through the body and may damage the nearby healthy tissue unless carefully planned. Radiation treatment planning is the process of finding the best system parameters to deliver sufficient dose to target structures while avoiding damage to nearby healthy tissue. This thesis investigates optimization approaches for planning external beam radiotherapy.

Modern external beam radiation systems contain a linear accelerator (LINAC), which can rotate around the patient who is positioned in a treatment couch as shown in Figure 1. The beams from the accelerator are shaped by a computer-controlled multi-leaf collimator (MLC) that is integrated to LINAC as shown in Figure 2. The MLC contains parallel metallic leaf pairs which can be controlled individually to block some parts of the field. The open field formed by the MLC is called an *aperture* and the amount of energy delivered from the aperture per unit volume of tissue is called its *intensity*. Each aperture shape (beam shape) is modeled as a collection of pencil beams (called beamlets) and has a uniform intensity. A desired intensity pattern can be obtained by the additive effect of a series

Figure 1: A Linear Accelerator

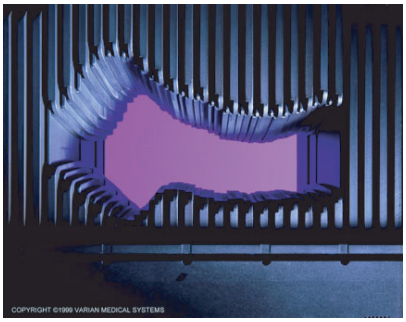

Figure 2: Multi-Leaf Collimators are Used to Create Beam Shapes

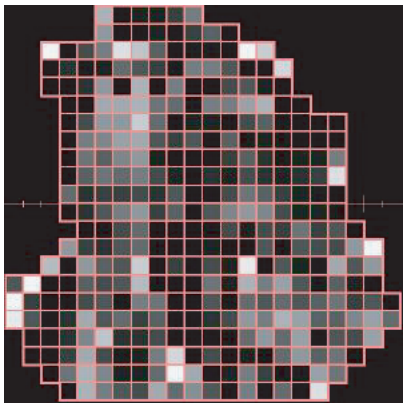

Figure 3: An Intensity Pattern Formed by a Series of Beam Shapes

of aperture shapes and intensities. Figure 3 shows an intensity pattern formed by the superposition of multiple beam shapes from a fixed direction. The beamlets (pixels) are colored based on their radiation intensities.

In this thesis, we focus on treatment planning approaches for Intensity Modulated Radiation Therapy (IMRT) and Volumetric Modulated Arc Therapy which will be described next. IMRT is a radiation treatment technique with multiple beams in which at least some of the beams are intensity-modulated and intentionally deliver a non-uniform intensity [5]. The radiation source is rotated around the treatment couch and is stopped at several positions to cross-fire at a cancerous tumor volume. These delivery positions are called *beam angles*. A selected set of beam angles for delivery is called a *beam geometry*. At each beam angle the relative position of the patient and the machine needs to be checked for accuracy. Therefore, the number of beam angles is limited (between 5 and 8) in order to reduce patient positioning times and chances for patient positioning errors. For many types of cancer such as prostate cancer, and head and neck cancer, the use of IMRT was reported to produce a highly concentrated treatment of the tumor volume, while limiting the radiation dose to adjacent healthy tissue. However, long setup times at each beam angle which decreases patient comfort and its limitation to utilize only a part of the coplanar space around the patient led to the development of a new delivery technique called Volumetric-Modulated Arc Therapy (VMAT). (The technology is called VMAT by Elekta Inc. and RapidArc by Varian Medical Systems; we will use VMAT throughout this thesis.) Treatment planning

technology for VMAT is still considered to be in its early stages of development, but the infrastructure has been widely adapted as Elekta reports more than 200 orders in three years since its introduction. Radiation is continuously delivered in VMAT during a single rotation of the treatment machine around the patient, while in IMRT delivery is made only at specific angles. The aperture shapes and the intensity levels are continuously modulated during the rotation. The main advantage of the VMAT technology is that the delivery is very fast, usually in less than 2 minutes compared to the treatment of the same disease site in 14 minutes with IMRT.

The patient anatomy may change significantly during a treatment session. This is especially true for lung cancers where the breathing cycle of the patient continuously changes the patient anatomy near the tumor. To handle anatomical changes, radiation is currently only delivered at one phase of the breathing cycle. However, that increases treatment time, so investigations into delivery during multiple phases of the breathing cycle are ongoing. Due to the inclusion of the time aspect resulting techniques are called Four-Dimensional (4D). We investigate 4D approaches in IMRT treatment planning.

The goal in radiation treatment planning is to find the best system parameters so that the delivered radiation will destroy or reduce the growth of cancer cells with minimal impact on nearby healthy organs. In Figure 4, we illustrate the typical work flow in a commercial planning system. Green colored boxes show processes that are handled automatically by a clinical software or hardware, and blue boxes indicate processes that require a human treatment planner. Orange boxes show the processes that require physicians. Computed tomography (CT) scans of the patient are used to contour target structures and healthy organs by a human planner. For modeling purposes, the structures (target and critical structures) are discretized into cubes called voxels (e.g., cubes of $5 \times 5 \times 5$ mm) and the dose delivered to each voxel per unit intensity of each beamlet is calculated. The total dose received by a voxel is the sum of doses deposited from each beamlet. The resulting data is called a *dose matrix* and is calculated by a clinical software. In order to guide the construction of a treatment plan, a radiation oncologist specifies a set of requirements that have to be satisfied in any acceptable treatment plan. These requirements are in the form of

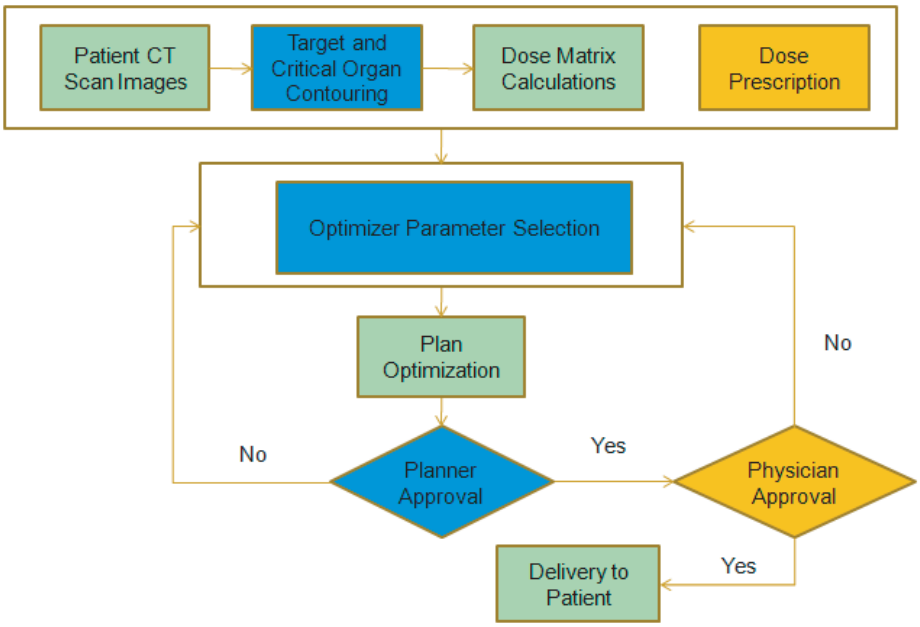

Figure 4: Typical Workflow in Treatment Planning

a minimum prescription dose for target structure voxels and a maximum tolerance dose for nearby critical structure voxels. A prescription dose is the dose level necessary to destroy or damage target cells, while a tolerance dose is the level above which complications for healthy tissues may occur. This part of the process can be thought of as the input stage for plan optimization. Next, a human planner picks the parameters that are required as input to a treatment plan optimization software. These parameters are mostly the weights assigned to different objectives that will guide the optimization algorithms to evaluate trade-offs which arise in balancing target radiation and healthy tissue sparing. After an expected time of 15-20 minutes the treatment planner checks the output from the optimization software. If he is satisfied, he passes the solution to the physician and after physician's approval the treatment plan is delivered to patient. However, if the treatment planner is not satisfied due to a possible violation of some prescription requirements he returns back to the parameter selection phase and reoptimizes the plan using the commercial planner. Sometimes, the solutions provided by the planner may not be approved by the physician and the planner once more returns back to the parameter selection phase. This iterative part of the process takes the longest time in treatment planning as it requires a significant amount of human

interaction. One of the key motivations of our research is the automation of the planning process to reduce human interaction, which will not only decrease planning costs, but also will lead to a better capacity utilization and treatment of more patients.

In addition to the set of dose requirements defining acceptable treatment plans, treatment planners and clinical oncologists use 2-dimensional dose volume histograms, dose color wash on CT scans (e.g. red color for regions receiving the highest dose and purple for regions receiving the lowest dose) and some evaluation metrics to assess the quality of a treatment plan. The evaluation of a treatment plan is complicated due to the fact that the requirements and their relative importance are subjective, as are the underlying trade-offs they are trying to capture. Use of evaluation metrics for plan evaluation is a critical part in the development of an automated planning system as they can be calculated easily by a computer and compared against their ideal values. In this thesis we consider the following evaluation metrics: cold spot, hot spot, coverage and conformity. These metrics are defined next. The following notation is used throughout. Let N denote the set of beamlets, S the set of structures, V_s the number of voxels in a structure $s \in S$, and z_{js} the dose received by voxel j of structure s. (Of course z_{js} depends on the treatment plan.) For simplicity of exposition we assume that there is only one target structure $\tau \in S$.

The *cold spot* of target structure τ is defined as the ratio of the minimum dose received by any of the voxels of the structure to the prescription dose of structure τ, i.e.,

$$\text{cold spot}(\tau) = \frac{\min\{z_{j\tau} : \ j = 1, \ldots, V_\tau\}}{PD_\tau}, \tag{1}$$

where PD_τ is the prescription dose for target structure. Similarly, the *hot spot* of target structure is defined as the ratio of the maximum dose received by any of the voxels of the structure to the prescription dose, i.e.,

$$\text{hot spot}(\tau) = \frac{\max\{z_{j\tau} : \ j = 1, \ldots, V_\tau\}}{PD_\tau}. \tag{2}$$

Ideally, every voxel in a target structure receives exactly the prescription dose. The cold spot metric and hot spot metric measure the deviation from this ideal situation by examining the voxel receiving the smallest dose and the voxel receiving the largest dose.

The *coverage* of target structure τ is the proportion of voxels receiving a dose greater than or equal to the prescription dose PD_τ, i.e.,

$$\text{coverage}(\tau) = \frac{\sum_{j=1}^{V_\tau} \mathbb{I}_+(z_{j\tau} - PD_\tau)}{V_\tau}, \tag{3}$$

where $\mathbb{I}_+$ is the indicator function for the non-negative real line, i.e., $\mathbb{I}_+(a)$ is equal to 1 if $a \geq 0$ and 0 otherwise. Note that $0 \leq \text{coverage}(\tau) \leq 1$ and that values closer to 1 are preferable. The *conformity* of target structure τ is the ratio of the number of voxels in the structure and its surrounding tissue receiving a dose greater than or equal to the prescription dose PD_τ to the number of voxels in the structure itself receiving greater than or equal to the prescription dose, i.e.,

$$\text{conformity}(\tau) = \frac{\sum_{s \in S} \sum_{j=1}^{V_s} \mathbb{I}_+(z_{js} - PD_\tau)}{\sum_{j=1}^{V_\tau} \mathbb{I}_+(z_{j\tau} - PD_\tau)}. \tag{4}$$

Note that $1 \leq \text{conformity}(\tau)$ and that values closer to 1 are preferable. The coverage and conformity metrics are illustrated in Figure 5. In schematic (a), the set of voxels receiving a dose greater than or equal to the prescription dose is exactly the set of voxels in target structure τ and thus $\text{coverage}(\tau) = \text{conformity}(\tau) = 1$; this is the ideal solution. In schematic (b) the set of voxels receiving a dose greater than or equal to the prescription dose is about twice the size of the set of voxels in target structure τ (and includes all the voxels of τ) and thus $\text{coverage}(\tau) = 1$ and $\text{conformity}(\tau) = 1.96$. In schematic (c) the set of voxels receiving a dose greater than or equal to the prescription dose is less than half of the size of the target structure τ (but located inside the target structure) and thus $\text{coverage}(\tau) = 0.36$ and $\text{conformity}(\tau) = 1$. Finally, in schematic (d) the set of voxels receiving a dose greater than or equal to the prescription dose has the same size as that of the target structure but is offset from the target. In this case $\text{coverage}(\tau) = 0.39$ and $\text{conformity}(\tau) = 2.56$. Due to the fact that coverage and conformity consider the target as a whole, these metrics are typically of higher importance than the cold spot and hot spot metrics. (See Lee et al. [20] and [21] for earlier use of these metrics for plan evaluation.)

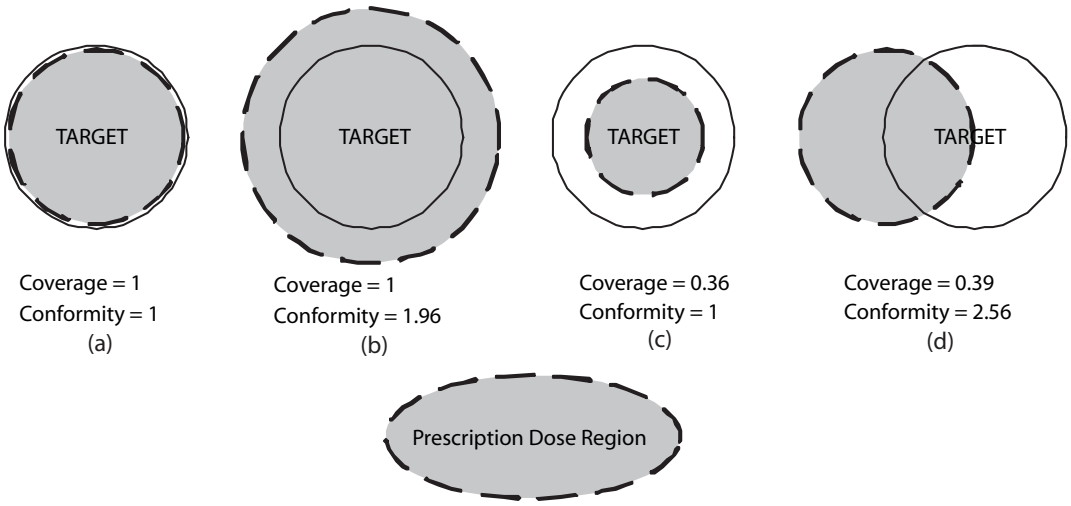

Figure 5: Coverage and Conformity Metrics

This thesis contributes to radiation treatment planning in several ways. First of all, we propose an alternative process as illustrated in Figure 6 for IMRT which is an automated treatment plan generation technology that can efficiently provide multiple high-quality plans for physician analysis and selection. Our technology addresses both the beam geometry and intensity aspects of the planning process and does not require human intervention for parameter selection. Our proposed approach is based on an existing model (Romeijn et al. [34]) for optimizing beamlet intensities (also called the fluence map) that uses conditional value-at-risk constraints (C-VaR constraints) to approximate constraints on the dose volume distribution. The use of C-VaR constraints has significant computational advantages as they can be handled using linear programming (LP) models, which are efficiently solved. However, the parameters controlling the C-VaR constraints have to be chosen carefully to get an accurate approximation of the dose-volume constraints.

We first show how the parameters of the C-VaR constraints can be used to control the coverage and conformity measures of plan quality. This is achieved by introducing C-VaR constraints on the target tumor volume as well as *virtual critical structures*. Virtual critical structures surround target volumes and are implemented to control the dose deposits specifically at the boundary of the target volume. Next we propose and implement a bicriteria parameter-tuning strategy for automatically generating multiple high-quality treatment plans. Key to this approach is the possibility of efficiently solving the underlying

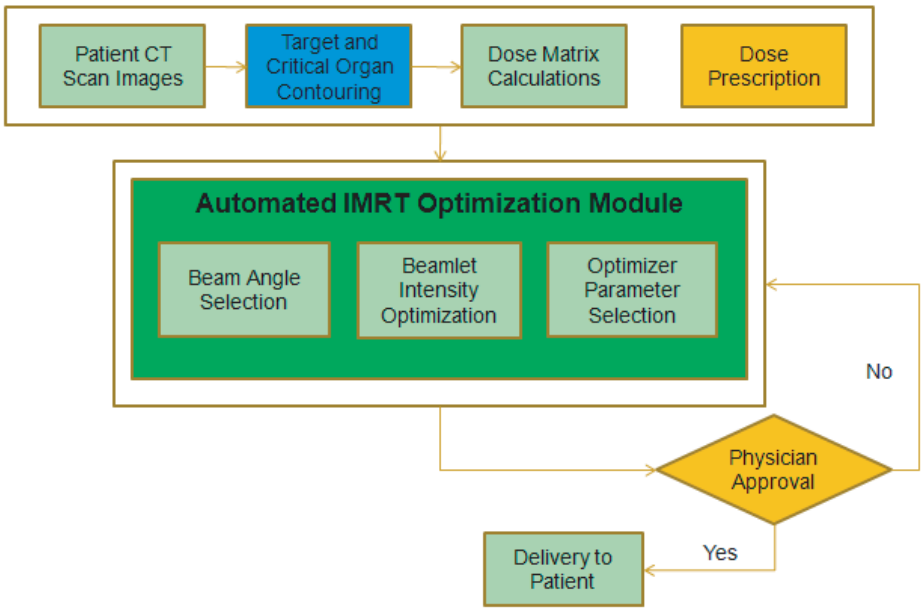

Figure 6: Proposed Process for IMRT Planning

LP-based fluence map optimization (FMO) problem. Finally we use the above mentioned FMO model within a heuristic for selecting beam geometries. At the heart of the scheme is a multi-attribute beam scoring mechanism based on a treatment plan constructed for a beam geometry involving a large number of beam angles.

The beam selection algorithm produces better plans compared to a well-known scoring-based method by Pugachev and Lei [33] in computational studies on three different tumor sites: brain, head and neck, prostate. The algorithm finds the best beam configuration at least ten times faster in all test cases as compared to an exact mixed integer programming model. The algorithm is able to produce similar or higher quality solutions with less number of beams compared to equi-spaced beam configurations. The improvements are clearly observed in the values of evaluation metrics, dose-volume histograms and colorwash dose distributions overlaid with CT images. Solution times range from 2 minutes to 15 minutes which are clinically acceptable.

The changes in patient anatomy during a treatment session due to breathing needs to be considered in treatment planning especially for lung cancer. Currently, a stable phase of the breathing cycle is selected for dose delivery to the patient and the radiation is shut-off during

other phases (single-phase gating). We investigate an alternative treatment mode where the radiation is allowed to be delivered in multiple phases. We find optimal fluence maps for each phase of the breathing cycle simultaneously by considering the overall dose delivered to patient; the optimization model decides how much dose to deliver in each phase. Because the optimization exploits the opportunities provided in each phase better treatment plans can be obtained. We investigate the benefit of this approach with computational studies on a real lung case patient and show that the tumor coverage can be improved from only 51% in single-phase gating to 96% with a five-phase gating solution.

To be able to calculate the total dose received by each voxel at the end of treatment, it is necessary that the dose received by a voxel of a structure per intensity of a beamlet is available for each phase and that each voxel can be tracked through the different phases. This implies that we assume sufficiently powerful image registration algorithms are available to do so. The treatment plan will specify the intensity of each beamlet in each phase and the dose received by each voxel of a structure at the end of treatment. Under these assumptions it is possible to extend IMRT formulations to optimize the total dose received by the patient over all phases. The resulting treatment is quite different from single-phase gating as the patient is treated during multiple phases; the optimization model decides how much dose to deliver in each phase. This reduces the treatment time and thus the comfort level of the patient. Furthermore, because the optimization exploits the opportunities provided in each phase better treatment plans can be obtained.

VMAT has the potential to provide treatments in less time compared to other delivery techniques which enhances patient comfort and allows for the treatment of more patients. The planning system decides how the beam shapes and dose rates should change during the rotation of the radiation source. In this thesis, we provide two different planning approaches for VMAT. The first approach tries to find leaf positions and dose rates separately in a two-stage algorithm. First, the leaf positions are selected by assigning scores to individual beamlets using their geometric and dosimetric properties, and maximizing the total score using a shortest-path algorithm in order to keep the favorable beamlets open. Once the leaf positions are selected, a linear program is solved to optimize dose rates. Although

this approach can produce treatment plans very efficiently the results are not of acceptable quality for clinical implementation. The second approach tries to optimize leaf positions and dose rates simultaneously by constructing a large scale mixed-integer programming model. The solution of this model with all decision alternatives is computationally intractable considering the clinical time limits. Therefore, we develop a heuristic algorithm which solves this model multiple times on a reduced set of of decision variables by identifying candidate leaf positions and control points at each iteration. We develop valid inequalities for the model which not only decrease the solution times but also allows us to get better integer solutions in specified time limits. Computational studies on a spinal tumor and a prostate tumor case produce clinically acceptable results.

The remainder of the thesis is organized as follows. In Chapter 2, we describe our proposed IMRT planning system. In Chapter 3, the IMRT algorithms are extended to a multi-phase 4D-IMRT planning method to show how such multi-phase solutions can improve plan quality for lung cases where breathing causes significant anatomical changes during the treatment. In Chapter 4, we propose planning algorithms for VMAT which is expected to succeed IMRT in the future due to its important potential advantage on reducing treatment times while providing similar quality results.

CHAPTER II

AN AUTOMATED INTENSITY-MODULATED RADIATION THERAPY PLANNING SYSTEM

External beam radiation therapy is used to treat over 500,000 cancer patients annually in the United States. This therapy uses multiple beams of radiation from different directions to cross-fire at a cancerous tumor volume in order to kill the cancer cells, thereby shrinking the tumor. Intensity-modulated radiation therapy (IMRT) is an advanced mode of high-precision external radiotherapy that utilizes computer-controlled mega-voltage x-ray accelerators to deliver precise radiation doses to targeted tissues. Rather than being treated with a large uniform beam, in IMRT the patient is treated by a series of beam shapes. Each shape is modeled as a collection of pencil beams (called beamlets). For many types of cancer, such as prostate cancer and head and neck cancer, the use of intensity modulation allows a highly concentrated treatment of the tumor volume, while limiting the radiation dose to adjacent healthy tissue (see Veldeman et al. [46] for comparisons of IMRT and non-IMRT treatments on different tumor sites).

Constructing an IMRT treatment plan that radiates the cancerous tumor volume (called target) without impacting adjacent normal structures (called organs at risk) is challenging. The planning is concerned with selecting a beam geometry and beamlet intensities to produce the best dose distribution that can be delivered efficiently. Because of the many possible beam geometries and the range of intensities, there is an infinite number of treatment plans, and consistently and efficiently generating high-quality treatment plans is complicated. A significant challenge is that there is no single metric to assess the quality of a treatment plan; therefore, trade-offs have to be made. Typically, a radiation oncologist specifies a set of requirements (a prescription) that has to be satisfied in any acceptable treatment plan. A number of measures have been developed to assess quality of acceptable treatment plans, e.g. the coverage of a target volume by a prescription dose, the conformity

of a prescription dose around a target volume, and the highest and the lowest doses received by a target volume (Lee et al. [21]). The oncologist also considers dose-volume histograms (DVHs) depicting the dose distributions over the structures (both target volumes and organs at risk) associated with the treatment. The quality measures and DVHs are used to choose among acceptable treatment plans. Further complicating the evaluation of a treatment plan is the fact that the requirements and their relative importance are subjective, as are the underlying trade-offs they are trying to capture.

A variety of optimization based approaches have been developed for the different aspects (determining beam geometries, intensities, and delivery options) of the IMRT planning process. These approaches are iterative in nature and necessitate human evaluation and guidance. This makes the process time consuming and costly. In this chapter we propose an automated treatment plan generation technology that can efficiently provide multiple high-quality plans for physician analysis and selection. Our technology addresses both the beam geometry and intensity aspects of the planning process. The final stage of the IMRT planning problem in which the calculated beamlet intensities are converted to a series of beam shapes for efficient delivery is not addressed in this chapter. For a more detailed description, review, history, and physical basis of IMRT, see Bortfeld [5], Boyer et al. [6], Shepard et al. [40] and Webb [49].

Our proposed approach is based on an existing model (as a result of Romeijn et al. [34]) for optimizing beamlet intensities (also called the fluence map) that uses conditional value-at-risk constraints (C-VaR constraints) to approximate constraints on the dose volume distribution. The use of C-VaR constraints has significant computational advantages as they can be handled using linear programming (LP) models, which are efficiently solved. However, the parameters controlling the C-VaR constraints have to be chosen carefully to get an accurate approximation of the dose-volume constraints.

We first show how the parameters of the C-VaR constraints can be used to control the coverage and conformity measures of plan quality. This is achieved by introducing C-VaR constraints on the target tumor volume as well as *virtual critical structures*. Virtual critical structures surround target volumes and are implemented to control the dose deposits

specifically at the boundary of the target volume (Bahr et al. [2], Price et al. [32], Lee et al. [21]). Next we propose and implement a bicriteria parameter-tuning strategy for automatically generating multiple high-quality treatment plans. Key to this approach is the possibility of efficiently solving the underlying LP-based fluence map optimization (FMO) problem. Finally we use the above mentioned FMO model within a heuristic for selecting beam geometries. At the heart of the scheme is a multi-attribute beam scoring mechanism based on a treatment plan constructed for a beam geometry involving a large number of beam angles.

In summary, we develop a treatment plan generation technology that optimizes both beam geometry and beamlet intensities. The technology is automated and generates several high-quality treatment plans satisfying the provided requirements in a single invocation and without human guidance. The technology has been tested on various real patient cases with success. Solution times range from a few minutes to a quarter of an hour which are clinically acceptable.

The remainder of the the chapter is organized as follows. In Section 2.1, we provide details on the IMRT treatment planning problem and the various evaluation metrics considered in this work. In Section 2.2, we describe the core components of our IMRT treatment plan generation technology. Finally, in Section 2.3, we present the results of an extensive computational study.

2.1 Problem Description

An IMRT treatment plan has to specify a beam geometry (a set of beam angles) and for each beam angle a fluence map (a set of beamlet intensities; a beam can be thought of as consisting of a number of beamlets that can be controlled individually). Typically, a small number of equi-spaced *coplanar* beam angles are used. Coplanar beam angles are obtained by rotating the gantry while keeping the treatment couch in a fixed position parallel to the gantry axis of rotation. There are practical reasons for limiting the number of beam angles (between 5 and 8) as it reduces patient positioning times, chances for patient positioning errors, and delivery time.

For modeling purposes, the structures (target and critical structures) are discretized into cubes called voxels (e.g., cubes of $5 \times 5 \times 5$ mm) and the dose delivered to each voxel per intensity of each beamlet is calculated (see Ahnesjo [1], Mackie et al. [27], and Mohan et al. [28] for dose calculation methods). The total dose received by a voxel is the sum of doses deposited from each beamlet. The goal is to build a treatment plan that creates a dose distribution that will destroy, or at least damage, target cells while sparing healthy tissue by choosing proper beam angles and beamlet intensities.

In order to guide the construction of a treatment plan, a radiation oncologist specifies a set of requirements that have to be satisfied in any acceptable treatment plan. These requirements are in the form of a minimum prescription dose for target structure voxels and a maximum tolerance dose for nearby critical structure voxels. A prescription dose is the dose level necessary to destroy or damage target cells, while a tolerance dose is the level above which complications for healthy tissues may occur.

In addition to the set of dose requirements defining acceptable treatment plans, clinical oncologists use a set of evaluation metrics to assess the quality of a treatment plan. In this chapter we consider the following metrics: cold spot, hot spot, coverage and conformity. These metrics are defined next. The following notation is used throughout. Let N denote the set of beamlets, S the set of structures, V_s the number of voxels in a structure $s \in S$, and z_{js} the dose received by voxel j of structure s. (Of course z_{js} depends on the treatment plan.) For simplicity of exposition we assume that there is only one target structure $\tau \in S$.

The *cold spot* of target structure τ is defined as the ratio of the minimum dose received by any of the voxels of the structure to the prescription dose of structure τ, i.e.,

$$\text{cold spot}(\tau) = \frac{\min\{z_{j\tau} : \ j = 1, \ldots, V_\tau\}}{PD_\tau}, \tag{5}$$

where PD_τ is the prescription dose for target structure. Similarly, the *hot spot* of target structure is defined as the ratio of the maximum dose received by any of the voxels of the structure to the prescription dose, i.e.,

$$\text{hot spot}(\tau) = \frac{\max\{z_{j\tau} : \ j = 1, \ldots, V_\tau\}}{PD_\tau}. \tag{6}$$

Ideally, every voxel in a target structure receives exactly the prescription dose. The cold spot

metric and hot spot metric measure the deviation from this ideal situation by examining the voxel receiving the smallest dose and the voxel receiving the largest dose.

The *coverage* of target structure τ is the proportion of voxels receiving a dose greater than or equal to the prescription dose PD_τ, i.e.,

$$\text{coverage}(\tau) = \frac{\sum_{j=1}^{V_\tau} \mathbb{I}_+(z_{j\tau} - PD_\tau)}{V_\tau}, \tag{7}$$

where $\mathbb{I}_+$ is the indicator function for the non-negative real line, i.e., $\mathbb{I}_+(a)$ is equal to 1 if $a \geq 0$ and 0 otherwise. Note that $0 \leq \text{coverage}(\tau) \leq 1$ and that values closer to 1 are preferable. The *conformity* of target structure τ is the ratio of the number of voxels in the structure and its surrounding tissue receiving a dose greater than or equal to the prescription dose PD_τ to the number of voxels in the structure itself receiving greater than or equal to the prescription dose, i.e.,

$$\text{conformity}(\tau) = \frac{\sum_{s \in S} \sum_{j=1}^{V_s} \mathbb{I}_+(z_{js} - PD_\tau)}{\sum_{j=1}^{V_\tau} \mathbb{I}_+(z_{j\tau} - PD_\tau)}. \tag{8}$$

Note that $1 \leq \text{conformity}(\tau)$ and that values closer to 1 are preferable. The coverage and conformity metrics are illustrated in Figure 7. In schematic (a), the set of voxels receiving a dose greater than or equal to the prescription dose is exactly the set of voxels in target structure τ and thus $\text{coverage}(\tau) = \text{conformity}(\tau) = 1$; this is the ideal solution. In schematic (b) the set of voxels receiving a dose greater than or equal to the prescription dose is about twice the size of the set of voxels in target structure τ (and includes all the voxels of τ) and thus $\text{coverage}(\tau) = 1$ and $\text{conformity}(\tau) = 1.96$. In schematic (c) the set of voxels receiving a dose greater than or equal to the prescription dose is less than half of the size of the target structure τ (but located inside the target structure) and thus $\text{coverage}(\tau) = 0.36$ and $\text{conformity}(\tau) = 1$. Finally, in schematic (d) the set of voxels receiving a dose greater than or equal to the prescription dose has the same size as that of the target structure but is offset from the target. In this case $\text{coverage}(\tau) = 0.39$ and $\text{conformity}(\tau) = 2.56$. Due to the fact that coverage and conformity consider the target as a whole, these metrics are

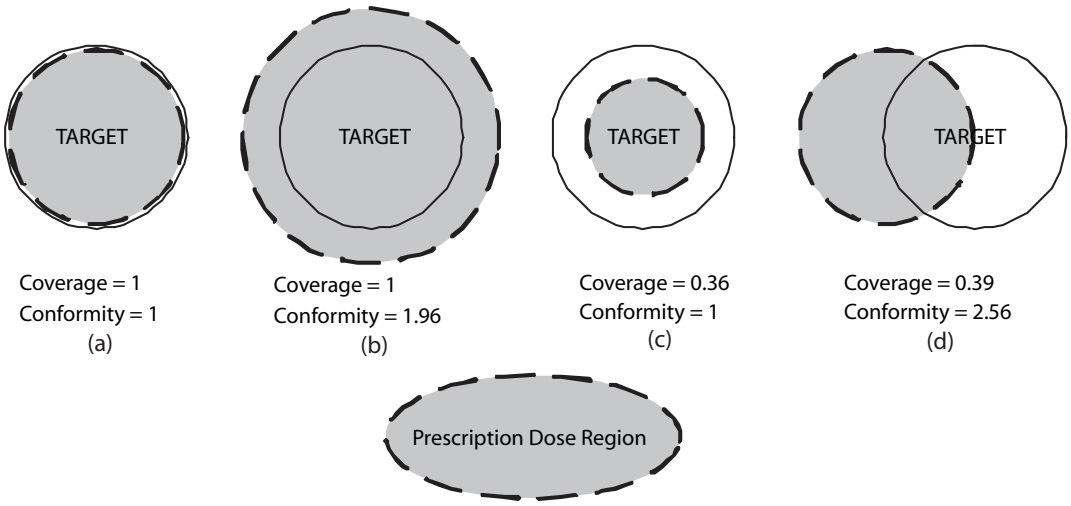

Figure 7: Coverage and Conformity Metrics

typically of higher importance than the cold spot and hot spot metrics. (See Lee et al. [20] and [21] for earlier use of these metrics for plan evaluation.)

The goal of our proposed approach is to determine fluence map and beam geometry consistent with treatment plans that satisfy the clinical dose requirements and have ideal values for the evaluation metrics. In the next section, we present an optimization-based approach towards this end.

2.2 Methodology

We first describe the C-VaR based FMO model of Romeijn et al. [34] and show how the C-VaR parameters can be used to alter coverage and conformity. Next, we describe a bicriteria search strategy for generating multiple high-quality treatment plans. Finally, beam angle selection is integrated with fluence map optimization. The beam angle selection scheme employs a bicriteria scoring of beam angle geometries and a selection mechanism to choose from among the set of non-dominated geometries.

2.2.1 Fluence Map Optimization Model

Let D_{ijs} be the dose received by voxel j of structure s per unit intensity of beamlet i. Let x_i be the intensity of beamlet i, i.e., a decision variable, then the dose z_{js} received by

voxel j of structure s, is

$$z_{js} = \sum_{i \in N} D_{ijs} x_i \qquad \forall j = 1, ..., V_s;\ s \in S. \tag{9}$$

2.2.1.1 Full volume constraints

Let L_s and U_s be the prescribed lower and upper dose limits for structure s, respectively. The full-volume constraints for dose limits are:

$$L_s \leq z_{js} \leq U_s \qquad \forall j = 1, ..., V_s;\ s \in S. \tag{10}$$

Prescribed requirements on cold spot and hot spot can be enforced by the bounds L_τ and U_τ in the full volume constraints (40).

2.2.1.2 Partial volume constraints

Partial-volume constraints specify dose limits that have to be satisfied by a specified fraction of the voxels of a structure. For target structure τ these constraints are formulated as the C-VaR constraint:

$$c_\tau - \frac{1}{(1-\alpha_\tau).V_\tau} \sum_{j=1}^{V_\tau} (c_\tau - z_{j\tau})^+ \geq L_\tau^\alpha \tag{11}$$

where $(a)^+$ is equal to a if $a \geq 0$ and 0 otherwise. This C-VaR constraint enforces that the average dose in the $(1-\alpha_\tau)$-fraction of voxels of target structure τ receiving the *lowest* amount of dose is *greater* than or equal to L_τ^α. If satisfied, at least $\alpha_\tau \times 100$ per cent of the voxels receive a dose greater than or equal to L_τ^α. For example, with $\alpha_\tau = 0.9$ and $L_\tau^\alpha = 30$ at least 90% of the voxels will receive a dose greater than 30 Gy. Here c_τ is a free variable associated with the dose-volume constraint of structure τ.

Similarly, the partial-volume constraints for critical structures are:

$$c_s + \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (z_{js} - c_s)^+ \leq U_s^\alpha \qquad s \in S. \tag{12}$$

For critical structure s the average dose in the subset of size $(1-\alpha_s)$ receiving the *highest* amount of dose is required to be *less* than or equal to U_s^α. When satisfied at least α_s percent of the voxels will receive a dose less than or equal to U_s^α.

Due to the indicator function in (7), requirements on coverage are difficult to model using convex constraints. Instead, we use partial volume constraints to enforce the coverage requirements using the following result.

Proposition 1 *Given $\alpha_s \in (0,1)$, if there exists a c_s such that*

$$c_s - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (c_s - z_{js})^+ \geq PD_s \tag{13}$$

then

$$\text{coverage}(s) \geq \alpha_s. \tag{14}$$

Proof. For simplicity of exposition, we assume that α_s is such that $\alpha_s V_s$ is a positive integer. Let D^{α_s} be such that

$$\sum_{j=1}^{V_s} \mathbb{I}_+(z_{js} - D^{\alpha_s}) = \alpha_s V_s.$$

Then the coverage requirement (14) is equivalent to

$$D^{\alpha_s} \geq PD_s.$$

We prove the result by showing that if there exists c_s satisfying (13) then we must have $D^{\alpha_s} \geq PD_s$.

Note that

$$D^{\alpha_s} \geq D^{\alpha_s} - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+,$$

which follows simply from the fact that $(D^{\alpha_s} - z_{js})^+$ is non-negative. We claim that

$$D^{\alpha_s} - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ \geq c_s - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (c_s - z_{js})^+ \quad \forall\ c_s \in \Re.$$

Suppose this is not true, that is, there exists c_s^* such that

$$D^{\alpha_s} - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ < c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (c_s^* - z_{js})^+, \tag{15}$$

and consider the following two cases.

Case 1: suppose that $c_s^* \leq D^{\alpha_s}$. Then the right-hand-side (rhs) of (15) is

$$
\begin{aligned}
\text{rhs} &= c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ c_s^* > z_{js}} (c_s^* - z_{js}) \\
&= c_s^* - \frac{1}{(1-\alpha_s).V_s} \left[\sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - z_{js}) - \sum_{j:\ c_s^* \le z_{js} < D^{\alpha_s}} (c_s^* - z_{js}) \right] \\
&= c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - z_{js}) + \underbrace{\frac{1}{(1-\alpha_s).V_s} \sum_{j:\ c_s^* \le z_{js} < D^{\alpha_s}} (c_s^* - z_{js})}_{\le 0} \\
&\le c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - D^{\alpha_s} + D^{\alpha_s} - z_{js}) \\
&= c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - D^{\alpha_s}) \\
&= c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ - (c_s^* - D^{\alpha_s}) \\
&= D^{\alpha_s} - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ = \text{lhs},
\end{aligned}
$$

where lhs is the left-hand-side of inequality (15), hence a contradiction. Note that the last step in the above follows from the definition of D^{α_s} which implies $|\{j:\ z_{js} < D^{\alpha_s}\}| = (1-\alpha_s)V_s$.

Case 2: suppose that $c_s^* > D^{\alpha_s}$. Then the right-hand-side (rhs) of (15) is

$$
\begin{aligned}
\text{rhs} &= c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ c_s^* > z_{js}} (c_s^* - z_{js}) \\
&= c_s^* - \frac{1}{(1-\alpha_s).V_s} \left[\sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - z_{js}) + \sum_{j:\ D^{\alpha_s} \le z_{js} < c_s^*} (c_s^* - z_{js}) \right] \\
&= c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - z_{js}) - \underbrace{\frac{1}{(1-\alpha_s).V_s} \sum_{j:\ D^{\alpha_s} \le z_{js} < c_s^*} (c_s^* - z_{js})}_{\ge 0} \\
&\le c_s^* - \frac{1}{(1-\alpha_s).V_s} \sum_{j:\ D^{\alpha_s} > z_{js}} (c_s^* - D^{\alpha_s} + D^{\alpha_s} - z_{js}) \\
&= D^{\alpha_s} - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ = \text{lhs},
\end{aligned}
$$

where the last step is identical to Case 1. Hence we again have a contradiction.

Thus if there exists c_s satisfying (13) then

$$D^{\alpha_s} \geq D^{\alpha_s} - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (D^{\alpha_s} - z_{js})^+ \geq c_s - \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (c_s - z_{js})^+ \geq PD_s$$

completing the proof. □

Thus coverage(τ) can be increased by increasing α_τ in the partial volume constraints (41) with $L_\tau^\alpha = PD_\tau$. Of course, $\alpha_\tau = 1$ is preferable, but such an aggressive value may lead to infeasibility. Note that if $\alpha_\tau = 1$ the partial volume constraint reduces to a full volume constraint.

Conformity requirements cannot be directly enforced by full or partial volume constraints on the dose on the target structure. This is because the conformity metric considers dose deposited outside the target. In order to keep track of the intended dose for target τ that is deposited outside the target, we use a *virtual critical structure* around the target structure τ. This is typically a 30 mm band around the target structure τ, denoted by structure $\beta \in S$. Bahr et al. [2], Price et al. [32] and Lee et al. [21] showed on clinical cases that limiting or minimizing the dose on virtual critical structures around the target structures can improve the conformity of the plans. Girinsky et al. [15] also used virtual critical structures to avoid high doses around the tumor volume.

Assuming that *all* the dose intended for the target structure τ is deposited in the voxels in structures τ and β, the only structure in S other than τ that can have more than the prescribed dose PD_τ will be structure β. Then from (7)and (8)

$$\begin{aligned} \text{conformity}(\tau) &= 1 + \frac{\sum_{j=1}^{V_\beta} \mathbb{I}_+(z_{j\beta} - PD_\tau)}{\sum_{j=1}^{V_\tau} \mathbb{I}_+(z_{j\tau} - PD_\tau)} \\ &= 1 + \frac{\text{coverage}(\beta) \cdot V_\beta}{\text{coverage}(\tau) \cdot V_\tau}, \end{aligned}$$

.

We know that by enforcing the partial volume constraint (41) on structure τ with $L_\tau^\alpha = PD_\tau$ for $\alpha_\tau \in (0,1)$, we ensure coverage(τ) $\geq \alpha_\tau$. Similarly, it can be shown that by

enforcing the partial volume constraint (42) on structure β with $U_\beta^\alpha = PD_\tau$ for $\alpha_\beta \in (0,1)$, we ensure coverage$(\beta) \leq (1-\alpha_\beta)$. Thus

$$\text{conformity}(\tau) \leq 1 + \frac{(1-\alpha_\beta)\cdot V_\beta}{\alpha_\tau \cdot V_\tau}. \tag{16}$$

Therefore we can reduce conformity(τ) by increasing α_τ and α_β. Setting these parameters at their maximum value of 1 typically leads to problem infeasibility, and so a careful selection is important. This is further discussed in Section 2.2.2.

Including the partial-volume constraints our FMO model is:

$$\min \quad \sum_{s\in S\backslash\{\tau\}} \frac{1}{V_s}\sum_{j=1}^{V_s} z_{js} - \frac{1}{V_\tau}\sum_{j=1}^{V_\tau} z_{j\tau} \tag{17}$$

$$\text{s.t.} \quad z_{js} = \sum_{i\in N} D_{ijs}x_i \quad \forall j = 1,...,V_s;\ s\in S \tag{18}$$

$$L_s \leq z_{js} \leq U_s \quad \forall j = 1,...,V_s; s\in S \tag{19}$$

$$c_s + \frac{1}{(1-\alpha_s)V_s}\sum_{j=1}^{V_s}(z_{js}-c_s)^+ \leq U_s^\alpha \quad s\in S\backslash\{\tau\} \tag{20}$$

$$c_\tau - \frac{1}{(1-\alpha_\tau)V_\tau}\sum_{j=1}^{V_\tau}(c_\tau - z_{j\tau})^+ \geq L_\tau^\alpha \tag{21}$$

$$x_i \geq 0 \quad i = 1,...,N \tag{22}$$

$$z_{js} \geq 0 \quad j = 1,...,V_s; s\in S \tag{23}$$

$$c_s \text{ free} \quad s\in S. \tag{24}$$

The objective function (32) attempts to decrease the average dose on the critical structures while increasing the average dose on the target structure. We can integrate various other objective functions, but we leave this as a future research subject (see Kessler et al. [19]; Yang and Xing [52] for objective functions used in IMRT formulations).

2.2.2 Parameter Search

As mentioned before, the coverage and conformity metrics for the target structures depend on the values of the parameters α_β and α_τ. In this section we describe a search technique to identify appropriate values of these parameters that improve treatment quality.

High values of α_β and α_τ are preferable for good coverage and conformity for the target

τ. However, this may lead to infeasibility. Let

$$F_\tau = \{(\alpha_\beta, \alpha_\tau) \in (0,1)^2 : \text{ The FMO model (32)-(39) is feasible}\}.$$

Note that F_τ is a monotone set, i.e., if $(\alpha_\tau^*, \alpha_\beta^*) \in F_\tau$, $\alpha_\tau \leq \alpha_\tau^*$ and $\alpha_\beta \leq \alpha_\beta^*$, then $(\alpha_\tau, \alpha_\beta) \in F_\tau$. Assuming F_τ is non-empty we would like to maximize α_τ and α_β over F_τ. Such solutions will lie on the upper boundary of F_τ. Our search technique starts with an initial solution and goes through several search phases to identify solutions on the upper boundary of F_τ. Figure 8 illustrates the search phases, which are detailed next.

Initial values of the parameters α_τ and α_β are chosen based on a minimum coverage requirement $MinCov$ and a maximum allowed conformity value $MaxConf$. In particular

$$\alpha_\tau^0 = MinCov \times \gamma \tag{25}$$

$$\alpha_\beta^0 = (1 - \frac{MinCov(MaxConf - 1)V_\tau}{V_\beta}) \times \gamma \tag{26}$$

where $0 < \gamma < 1$ is scaling factor to account for the fact that the bounds (14) and (16) are conservative. In our implementation for case studies we used $\gamma = 0.9$.

We solve the FMO model (32)-(39) with the initial parameter values $(\alpha_\beta^0, \alpha_\tau^0)$. If the problem is infeasible, i.e., $(\alpha_\beta^0, \alpha_\tau^0) \notin F_\tau$, then we decrease both parameters by λ, where $0 < \lambda < 1$, and resolve the FMO model. This is continued until a feasible set of parameters is found. The monotone structure of the set F_τ guarantees that if F_τ is non-empty, then Phase 0 will produce a feasible set of parameters.

Once we have initial feasible parameters, we execute Phase 1, where we increase both α_β and α_τ by λ and resolve the FMO model. This is repeated until we reach an infeasible set of parameters. The last set of feasible parameters, denoted by $(\alpha_\beta^1, \alpha_\tau^1)$, then is a solution near the upper boundary of F_τ.

Next we execute Phase 2 where, starting from $(\alpha_\beta^1, \alpha_\tau^1)$, we increase α_τ^1 by λ until we reach an infeasible set of parameters. The motivation here is that, since increasing α_τ improves both coverage and conformity, we want to obtain additional solutions with good coverage and conformity values. The last feasible solution from Phase 2 is denoted by $(\alpha_\beta^2, \alpha_\tau^2)$.

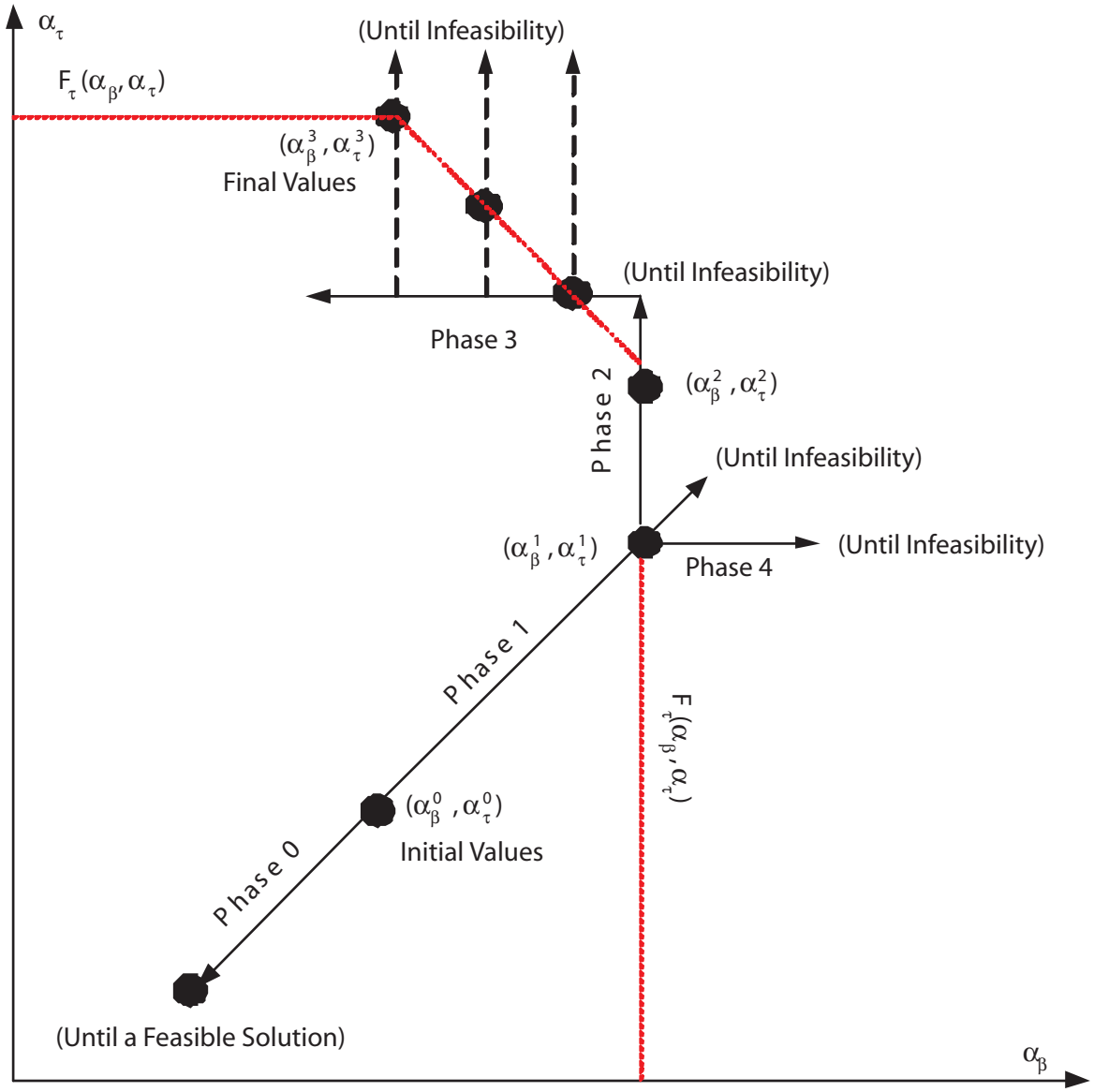

Figure 8: Parameter Search for α_τ and α_β

In order to find additional candidate solutions on the boundary of F_τ, we execute Phase 3. Here we reduce α_β by λ, and then start Phase 2 to increase α_τ. Note that in the solutions produced in this phase the conformity values may be worse than those from Phase 2, however coverage will not worsen.

Finally, if Phases 2 and 3 are not able to produce feasible parameters greater than $(\alpha_\beta^1, \alpha_\tau^1)$, then we execute Phase 4. Here we increase α_β by λ in order to find additional solutions near the upper boundary of F_τ.

The step size λ is an important component of the search process. From initial experimentations, we chose this to be $\lambda = 0.01$.

2.2.3 Beam Angle Selection

Careful selection of beam angles has long been recognized as an important part of creating IMRT plans (Stein et al. [42]). See Ehrgott et al. [13] for a rigorous presentation of the beam selection problem. Constructing an IMRT plan for all possible beam angle configurations and selecting the best is computationally prohibitive due to the number of possible configurations. Therefore, most algorithms assign scores to the candidate beam angles and use these scores to evaluate possible beam angle configurations. (See Bedford and Webb [4] for a classification of other beam selection algorithms proposed in the literature.) Scoring may be based on geometric and dosimetric information as in Pugachev and Lei [33]; or based on optimization of the fluence for each beam as in D'Souza et al. [11]. The drawback of these scoring-based approach is that the interaction between different beam angles is ignored. We introduce a bicriteria scoring approach to limit our choices to a relatively small number of non-dominated configurations. We then optimize the fluence only for these configurations to assess interaction effects. We compare the efficacy of our beam selection methodology with scoring-based approach of Pugachev and Lei [33] in Section 2.3.

We start by using the FMO model assuming that *all* candidate beam angles (say M) can be used. Let x_i^* be the intensity of beamlet i in beam (angle) B (where $B \in \{1, \ldots, M\}$) and $z_{j\tau}^*$ be the dose delivered to voxel j of the target structure τ in this treatment plan. Each beam B is then assigned the following two scores:

Total Dose Delivered to the Target Structure. The total dose delivered to the target structure by beam B is computed as

$$DPTV_B = \sum_{i \in B} \sum_{j=1}^{V_\tau} D_{ij\tau} x_i^*. \tag{27}$$

Preferring beams with a higher value of $DPTV_B$ favors beams that deliver higher dose to the target. The disadvantage of this measure is that it does not directly consider the dose received by target voxels in the lower tail of the dose distribution, i.e., the cold spots.

Weighted Sum of Dose Delivered to the Low Dose Region of the Target Structure. We define the low dose region C_τ of the target structure τ as the set of voxels that receive a dose less than 10% above the minimum dose observed in the target structure, i.e.,

$$C_\tau = \{j = 1, \ldots, V_\tau : \; z_{j\tau}^* \leq 1.10 \min_{j'}(z_{j'\tau}^*)\}. \tag{28}$$

The weighted sum of dose delivered to the low dose region of the target structure from beam angle B is

$$WPTV_B = \sum_{i \in B} \sum_{j \in C_\tau} D_{ij\tau} x_i^* \frac{1}{\max\{z_{j\tau}^*, \epsilon\}}, \tag{29}$$

where $\epsilon > 0$ is included to avoid division by zero. Note that we distinguish voxels in the low dose region by assigning weights inversely proportional to the actual dose received by these voxels. Preferring beams with higher value of $WPTV_B$ favors beams that deliver more dose to the low dose region of the target. The advantage of this measure is that it helps identify beams that deliver dose to what are commonly difficult regions to treat.

The score of a configuration with L angles selected from M candidates, i.e., a set of beam angles $C \subset \{1, \ldots, M\}$ such that $|C| = L$, is the sum of individual beam scores, i.e.,

$$DPTV(C) = \sum_{B \in C} DPTV_B, \text{ and}$$
$$WPTV(C) = \sum_{B \in C} WPTV_B.$$

We identify a set K of non-dominated configurations, i.e.,

$$K = \{C : \nexists\ D \text{ s.t. } DPTV(D) \geq DPTV(C) \text{ and } WPTV(D) > WPTV(C)\}\ \cup$$

$$\{C : \nexists\ D \text{ s.t. } DPTV(D) > DPTV(C) \text{ and } WPTV(D) \geq WPTV(C)\}.$$

The set K can be determined by enumerating all $\binom{M}{L}$ beam angle configurations. When the set of all configurations is very large, a more computationally effective approach is to solve a series of cardinality constrained knapsack problems to identify the non-dominated configurations. This scheme is illustrated in Algorithm 1.

Algorithm 1 Finding non-dominated beam angle configurations

Require: $DPTV_i$ and $WPTV_i$ scores of M candidate angles to choose L angles. Let $\epsilon > 0$ be small.

$RHS \leftarrow 0$

repeat

solve the cardinality constrained knapsack problem:

$$\begin{aligned} \max \quad & \sum_{i=1}^{M} DPTV_i y_i \\ \text{s.t.} \quad & \sum_{i=1}^{M} y_i = L \\ & \sum_{i=1}^{M} WPTV_i y_i \geq RHS \\ & y_i \in \{0,1\} \end{aligned}$$

if IP is feasible **then**

a non-dominated configuration is found

$RHS \leftarrow \sum_{i=1}^{M} WPTV_i y_i + \epsilon$

end if

until IP is infeasible

We would ideally generate an optimized treatment plan for each of the non-dominated configurations and select the one with the highest quality based on the evaluation metrics described earlier. However this is computationally intensive, and instead we perform a "greedy" search. We start with an arbitrary configuration $C \in K$ and calculate $(\alpha_\beta^2, \alpha_\tau^2)$. Next, we search for a configuration $\hat{C} \in K$ for which $(\alpha_\beta^2, \alpha_\tau^2 + \lambda)$ is feasible. If no such configuration $\hat{C}$ exists, then configuration C is selected. On the other hand, if such a

configuration $\hat{C}$ exists, we iterate and execute Phase 1 and Phase 2 of the parameter search for $\hat{C}$ starting from $(\alpha_\beta^2, \alpha_\tau^2)$.

2.3 Computational Study

In this section, we present the results of a computational study of three real-life cases: a pediatric brain case, a head and neck case, and a prostate case. The computational study focuses on the value of the techniques and solution schemes introduced above: (1) the development of an automated, systematic parameter search scheme for treatment plan construction models using C-VaR constraints to model partial dose volume constraints, and (2) the development of an efficient and effective beam angle selection scheme.

2.3.1 Case Descriptions

The three cases represent different parts of the body, and thus a variety of challenges in terms of treatment plan generation. So these cases demonstrate the robustness of our approach; no tuning is necessary for the specific cases.

Information concerning the prescription requirements for the target structures as well as for the critical structures in terms of full volume and partial volume dose constraints for the three cases can be found in Tables 1, 2, and 3. Partial dose volume constraints for the target structure (referred to as PTV) and the virtual critical structure (VCS) are computed using the formulas presented in Section 2.2 using a minimum coverage requirement of 0.95 and a maximum conformity requirement of 1.2. Observe that there are no partial dose-volume constraints for critical structures in the pediatric brain case.

To be able to analyze and judge the value of the different techniques and solution schemes, we start with base settings in each of the three cases. The base settings have 8 equi-spaced beams in the pediatric brain case, 8 equi-spaced beams in the head and neck case, and 6 equi-spaced beams in the prostate case. Furthermore, the full and partial dose volume constraints as specified in Tables 1, 2, and 3 are used.

Table 1: Brain Case - Structures and Constraints

Constraint Type	Structure	#Voxels	Percentage	L (cGy)	U (cGy)
Full volume	PTV	1,620	-	0	3,300
	Hypothalamus	88	-	0	2,160
	Right cochlea	26	-	0	1,260
	Left cochlea	21	-	0	1,260
	Pituitary	11	-	0	2,160
	Left eye	158	-	0	500
	Right eye	163	-	0	500
	VCS	6,581	-	0	3,300
Partial volume	PTV	1,620	95	3,060	-
	VCS	6,581	95	-	3,060

2.3.2 Value of the Parameter Search

As the treatment plan construction model only indirectly attempts to optimize coverage and conformity, we may be able to improve coverage and conformity by judiciously adjusting the dose-volume constraint parameters of the target structure and the virtual critical structure. We focus on the improvement in coverage and conformity that can be achieved through the parameter search. We compare the base settings (with full and partial dose volume constraints for the virtual critical structure included) with and without parameter search. The results are presented in Table 4. We observe that there are improvements for coverage and conformity for the brain and the prostate case. However, for the head-and-neck case the slight improvement in coverage is offset by a deterioration in conformity. The head and neck case also shows a significant cold spot. This is due to the fact that the target structure is up against the *Globe_RT* structure which has a full volume dose restriction of 2000 cGy. We will examine this issue in more detail in Section 2.3.6. The parameter search for the brain case and for the head and neck case end with a Phase 3 iteration, in which the virtual critical structure constraint is relaxed in hopes of finding a treatment plan with higher coverage. As a result, the value of conformity increases slightly in the last iteration. More details are provided in Figure 9 where we plot the change in the evaluation metrics during the course of the parameter search.

Table 2: Head and Neck Case - Structures and Constraints

Constraint Type	Structure	#Voxels	Percentage	L (cGy)	U (cGy)
Full volume	PTV	2,647	-	0	6,120
	Brainstem	1,075	-	0	5,400
	Spinal cord	387	-	0	4,500
	Globe_LT	388	-	0	2,000
	Globe_RT	355	-	0	2,000
	Optic_chiasm	13	-	0	5,400
	Optic_nerve_LT	26	-	0	5,400
	Optic_nerve_RT	10	-	0	5,400
	Parotid_LT	811	-	0	5,400
	Parotid_RT	832	-	0	5,400
	VCS	7,634	-	0	6,120
Partial volume	PTV	2,647	95	5,100	-
	Parotid_LT	811	50	-	2,600
	Parotid_RT	832	50	-	2,600
	VCS	7,634	93	-	5,100

2.3.3 Beam Angle Selection

A core ingredient of our angle selection scheme is identifying non-dominated beam configurations with respect to $DPTV_B$ and $WPTV_B$. In Figures 10, 11, and 12, we plot the scores for all 8-beam configurations for the brain and head and neck cases and all 6-beam configurations for the prostate case selected from 18 equi-spaced candidate beams and highlight the configurations that are non-dominated. The figures show that few non-dominated configurations exist. In Table 5, we provide additional information to support the observation that very few non-dominated configurations exist. For configurations of different sizes, we report the total number of configurations as well as the number of non-dominated configurations.

Next, we explore the effect of carefully selecting beam angles, and thus of integrating all the techniques and solution schemes developed. The results are presented in Table 6. The results are mostly self-explanatory and clearly demonstrate the value of the various techniques and solution schemes. However, these evaluation metrics only tell part of the story. Clinicians examine dose volume histograms and dose distributions to evaluate treatments, which are shown in Figures 13, 14, and 15 for the brain case, the head and neck case,

Table 3: Prostate Case - Structures and Constraints

Constraint Type	Structure	#Voxels	Percentage	L (cGy)	U (cGy)
Full volume	PTV	239	-	0	8694
	Rectum	1,267	-	0	8694
	Bladder	1,513	-	0	8694
	VCS	2,544	-	0	8694
Partial volume	PTV	239	95	7,560	-
	Rectum	2,544	70	-	7,560
	Rectum	2,544	50	-	4,500
	Bladder	1,513	50	-	4,500
	VCS	2,544	98	-	7,560

Table 4: Value of Parameter Search

	Brain		Head and Neck		Prostate	
	without parameter search	with parameter search	without parameter search	with parameter search	without parameter search	with parameter search
Coverage	0.957	0.984	0.956	0.959	0.950	1.000
Conformity	1.274	1.212	1.230	1.261	1.264	1.046
Coldspot	0.704	0.822	0.311	0.313	0.829	1.000
Hotspot	1.078	1.078	1.200	1.200	1.150	1.150

and the prostate case, respectively. Each of the figures displays dose volume histograms for three treatment plans: (1) a plan with equi-spaced beams obtained by solving the model, but without parameter search, (2) a plan with equi-spaced beams obtained by solving the model incorporating parameter search, and (3) a plan with optimized beam angles obtained by solving the model incorporating parameter search. The dose volume histograms show that improving coverage and conformity (i.e., by performing a parameter search) can have a negative impact on the dose volume histograms of the critical structures (increased doses delivered), but that by carefully selecting the beam angles these undesirable effects can be mostly negated. This is observed especially well in the head and neck case.

We show three sets of dose distribution images in Figures 16, 17, and 18 for the brain case, the head and neck case, and the prostate case, respectively. For each case, we present three slices corresponding to the three treatment plans: (1) a plan with equi-spaced beams

Table 5: Number of Non-dominated Configurations for Various Configuration Sizes

Configuration Size	# Non-Dominated Configurations			# Configurations
	Brain	Head and Neck	Prostate	
2	4	2	2	153
3	5	3	2	816
4	7	4	3	3,060
5	8	5	3	8,568
6	10	5	2	18,564
7	13	7	2	31,824
8	12	8	3	43,758
9	12	7	3	48,620

obtained by solving the model, but without parameter search, (2) a plan with equi-spaced beams obtained by solving the model incorporating parameter search, and (3) a plan with optimized beam angles obtained by solving the model incorporating parameter search. The prescription dose is shown as red colorwash with 75, 50 and 25% of prescription dose shown as yellow, green and blue colorwash. CT axial images 1, 2, and 3 represent inferior, mid and superior slices through the target volume. In all three cases, plan c (obtained by integrating parameter search with careful selection of beam angles) is preferable. In particular, in the brain case (Figure 16) the hot beamlet effect on the left side of the patient evident in plans a and b is minimized in plan c. In the head and neck case (Figure 17), for plan a, we observe significant 75% dose leaking (yellow colorwash) out of the tumor area in all three slices. On plan b, this is improved but we have a hot beamlet effect right next to the brainstem on slice 3 and this is not preferable. Plan c seems to be much better with tighter conformity of the dose including slice 3 where the PTV comes closer to brainstem. Similarly for the prostate case (Figure 18), plans a and b have hot spots along the beam directions which is not preferable. Plan c tends to pull in the high dose closer to the prostate.

The above results indicate that our beam angle selection scheme produces beam configurations that allow the construction of high-quality treatment plans. Next, we compare the proposed beam configurations to those produced by an optimization approach based on an integer programming formulation. We also compare our scheme to a scoring-based beam selection algorithm.

Table 6: Evaluation Metrics for Different Schemes

Case	Evaluation measure	Equi-spaced	Equi-spaced & VCS	Equi-spaced & VCS & Parameter search	Optimized angles & VCS & Parameter search
Brain	Coverage	0.954	0.957	0.984	1.000
	Conformity	1.533	1.274	1.212	1.125
	Coldspot	0.669	0.704	0.822	1.000
	Hotspot	1.078	1.078	1.078	1.078
Head and Neck	Coverage	0.953	0.956	0.959	0.974
	Conformity	1.742	1.230	1.261	1.232
	Coldspot	0.328	0.311	0.313	0.310
	Hotspot	1.200	1.200	1.200	1.200
Prostate	Coverage	0.933	0.950	1.000	1.000
	Conformity	1.601	1.264	1.046	1.033
	Coldspot	0.853	0.829	1.000	1.000
	Hotspot	1.150	1.150	1.150	1.150

Comparison with Mixed-Integer Programming Approach

We consider the MIP model for beam selection from Lee et al. [20, 21], Yang et al. [51], and Lim et al. [24]. We experimented with selecting 5 beam angles out of 8 equi-spaced candidate beam angles for the brain and the head and neck case; and selecting 6 beam angles out of 18 equi-spaced beam angles for the prostate case. (Due to the computational requirements of the optimization approach this comparison can only be performed for settings with a relatively small number of candidate beam angles for the brain and the head and neck cases.) For these cases, our selection scheme produced the optimal configuration, i.e., the same configuration that the optimization approach produced. It is informative to look at the difference in required computation time; see Table 7. It is clear that the optimization approach will become computationally prohibitive when 10 or more candidate beam angles are used.

Comparison with a Scoring-Based Beam Selection Algorithm

To further validate our approach, we compare our results to the results obtained by selecting the beam angles according to the pseudo Beam's-Eye-View (pBEV) scores introduced by Pugachev and Lei [33]. The steps for the pBEV calculation of a given beam angle are as

Table 7: Comparison with Mixed Integer Programming Approach

	Time (secs.)		Ratio
	Proposed Scheme	MIP	
Brain	348	5,676	1 : 16
Head and Neck	363	16,988	1 : 46
Prostate	52	540	1 : 10

follows:

1. Assign each beamlet an initial intensity value that delivers at least the prescription dose to every target voxel it is crossing;

2. For each critical structure voxel crossed by the beamlet, calculate the factor by which the initial beamlet intensity has to be multiplied to ensure the tolerance dose is not exceeded.

3. Find the minimum factor among all critical structures and adjust the initial beamlet intensity.

4. Perform a forward dose calculation using the beam intensity profile obtained;

5. Compute the score for chosen beam angle i according to an empiric score function as follows:

$$S_i = \frac{1}{V_s} \sum_{j=1}^{V_s} \left(\frac{d_{ij}}{PD_s}\right)^2 \tag{30}$$

where d_{ij} is the dose delivered to voxel j from beam angle i, V_s is the number of voxels in the target, and PD_s is the target prescription dose.

We select the final configuration simply by picking the highest scoring beams. (Pugachev and Lei [33] manually pick the beams in the case examples when some high-scoring angles are too close to each other to create a separation.)

We computed beam angle configurations for the cases using pBEV scores and our proposed scores, and then performed fluence map optimization. Multiple solutions are produced for each case by varying the number of beam angles. The results for brain case, head and neck case, and prostate case are presented in Tables 8, 9 and 10; respectively.

Table 8: Brain Case Comparison to pBEV

	6 beams		7 beams		8 beams		9 beams	
	pBEV	Proposed Scheme	pBEV	Proposed Scheme	pBEV	Proposed Scheme	pBEV	Proposed Scheme
Coverage	0.954	0.963	0.962	1.000	0.979	1.000	1.000	1.000
Conformity	1.245	1.214	1.279	1.144	1.233	1.125	1.082	1.099
Coldspot	0.919	0.956	0.928	1.000	0.947	1.000	1.000	1.000
Hotspot	1.078	1.078	1.078	1.078	1.078	1.078	1.078	1.078

Table 9: Head and Neck Case Comparison to pBEV

	6 beams		7 beams		8 beams		9 beams	
	pBEV	Proposed Scheme	pBEV	Proposed Scheme	pBEV	Proposed Scheme	pBEV	Proposed Scheme
Coverage	0.962	0.968	0.968	0.971	0.968	0.974	0.968	0.973
Conformity	1.253	1.219	1.247	1.243	1.199	1.232	1.177	1.162
Coldspot	0.337	0.301	0.327	0.320	0.351	0.310	0.360	0.314
Hotspot	1.200	1.200	1.200	1.200	1.200	1.200	1.200	1.200

The main difference between our scoring scheme and the pBEV scoring scheme is that we score beams based on their exact contributions to the delivered doses to target structures in an optimized solution rather than just using their geometric properties and prescription requirements. Another important difference is that we consider multiple beam configurations in the parameter search phase, which allows us to adjust dynamically in response to observed interaction effects between beams. In the pBEV algorithm there is only one configuration that is chosen based on initial parameter values.

2.3.4 Beam Configuration Size

The results presented so far assumed that the desired number of beam angles was decided in advance. We have seen that carefully selecting beam angles improves treatment plans. This suggests that it may be possible to get high-quality treatment plans with fewer beam angles. Figure 19 presents the values of the evaluation criteria for different sizes of beam configurations.

As expected, there are diminishing returns when we continue to increase the size of the configuration. This is especially clear for the prostate case where treatment plans of almost

Table 10: Prostate Case Comparison to pBEV

	4 beams		5 beams		6 beams	
	pBEV	Proposed Scheme	pBEV	Proposed Scheme	pBEV	Proposed Scheme
Coverage	1.000	1.000	1.000	1.000	1.000	1.000
Conformity	1.050	1.029	1.046	1.038	1.038	1.033
Coldspot	1.000	1.000	1.000	1.000	1.000	1.000
Hotspot	1.150	1.150	1.150	1.150	1.150	1.150

Table 11: Run Time (seconds) for Various Linear Programming Algorithms

	Voxels	Beamlets	#cons	#vars	primal simplex	dual simplex	barrier
Brain	8,668	1,010	17,881	16,871	1,548	1,395	48
Head and Neck	14,178	1,599	27,705	26,106	31,118	4,912	58
Prostate	5,563	1,146	12,398	13,544	377	35	24

equal quality are produced for configurations of size 4 and up. Furthermore, treatment plans with good quality can already be achieved with configurations of relatively small size (which has many practical advantages). If we compare the results presented in Table 6 to those presented in Figure 19, we see that carefully selecting beam angles allows us to construct treatment plans of equal or better quality than those that can be obtained with equi-spaced beams with fewer beams.

2.3.5 Solution Times

As mentioned above, we chose to develop only linear programming based technology, because linear programs can be solved efficiently. Several methodologies exist for solving linear programs, e.g., primal simplex, dual simplex, and interior point methods, and we examine the solution times for the different solution methods. We used the linear programming solver of XPRESS (Xpress-Optimizer 2007). Table 11 presents the instance characteristics for the head and neck case and the brain case (for 8 equi-spaced beams) and the prostate case (for 18 equi-spaced beams). All our experiments were run on a 2.66GHz Pentium Core 2 Duo processor with 3GB of RAM under Windows XP Operating System. The differences

are staggering; for this class of linear programs, interior point methods are by far superior as noted by Holder [16]. All remaining computational experiments therefore use XPRESS's barrier algorithm to solve linear programs.

The first step of our scheme to select beam angles is to solve the problem assuming that all candidate beam angles are used. In our computation experiments this meant solving the problem using 18 candidate beam angles (equi-spaced). The solution is used to get the beam angle scores $WPTV$ and $DPTV$. Table 12 presents the instance characteristics and the run times for the three cases. Once the beam angle scores are computed they are used

Table 12: Run Time for 18 Equi-Spaced Beams for Scoring

	Voxels	Beamlets	#cons	#vars	time (secs)
Brain	8,668	2,306	16,871	19,177	123
Head and Neck	14,178	3,650	26,106	29,756	278
Prostate	5,563	1,146	12,398	13,544	24

to identify non-dominated configurations. Then treatment plans are generated for these configurations (using a virtual critical structure and parameter search). In Table 13, we present the solution times for this component of the scheme for different beam configuration sizes. The maximum run time is about 20 minutes and observed for head and neck case to choose 9 beams. Our proposed scheme was able to produce high-quality solutions for the prostate case in less than 2 minutes.

2.3.6 Cold Spot Issue in the Head and Neck Case

In the head and neck case, we observed a significant cold spot in the treatment plan produced by our algorithm. Further analysis revealed that the cold spot results because

Table 13: Run Time (seconds) after Beam Scoring for Different Configuration Sizes

	2 beams	3 beams	4 beams	5 beams	6 beams	7 beams	8 beams	9 beams
Brain	145	199	263	514	478	689	798	445
Head and Neck	147	141	231	311	365	646	863	1,003
Prostate	28	12	14	23	28	35	50	53

the target structure (PTV) is up against the *Globe_RT* structure as can be seen in Figure 17 3.a, 3.b and 3.c. As the maximum tolerance dose for *Globe_RT* is 2,000 cGy and the target dose for the PTV is 5,100 cGy, it is not surprising that a cold spot results. The specified maximum tolerance dose of 2,000 cGy for *Globe_RT* is aggressive and upon seeing the generated treatment plan a clinician will likely increase the maximum tolerance dose. To explore what happens when the maximum tolerance dose for *Globe_RT* is increased, we present the evaluation metrics for the treatment plans generated for different maximum tolerance dose levels in Figure 20. We see that with a maximum tolerance dose of about 4,000 cGy the cold spot is eliminated.

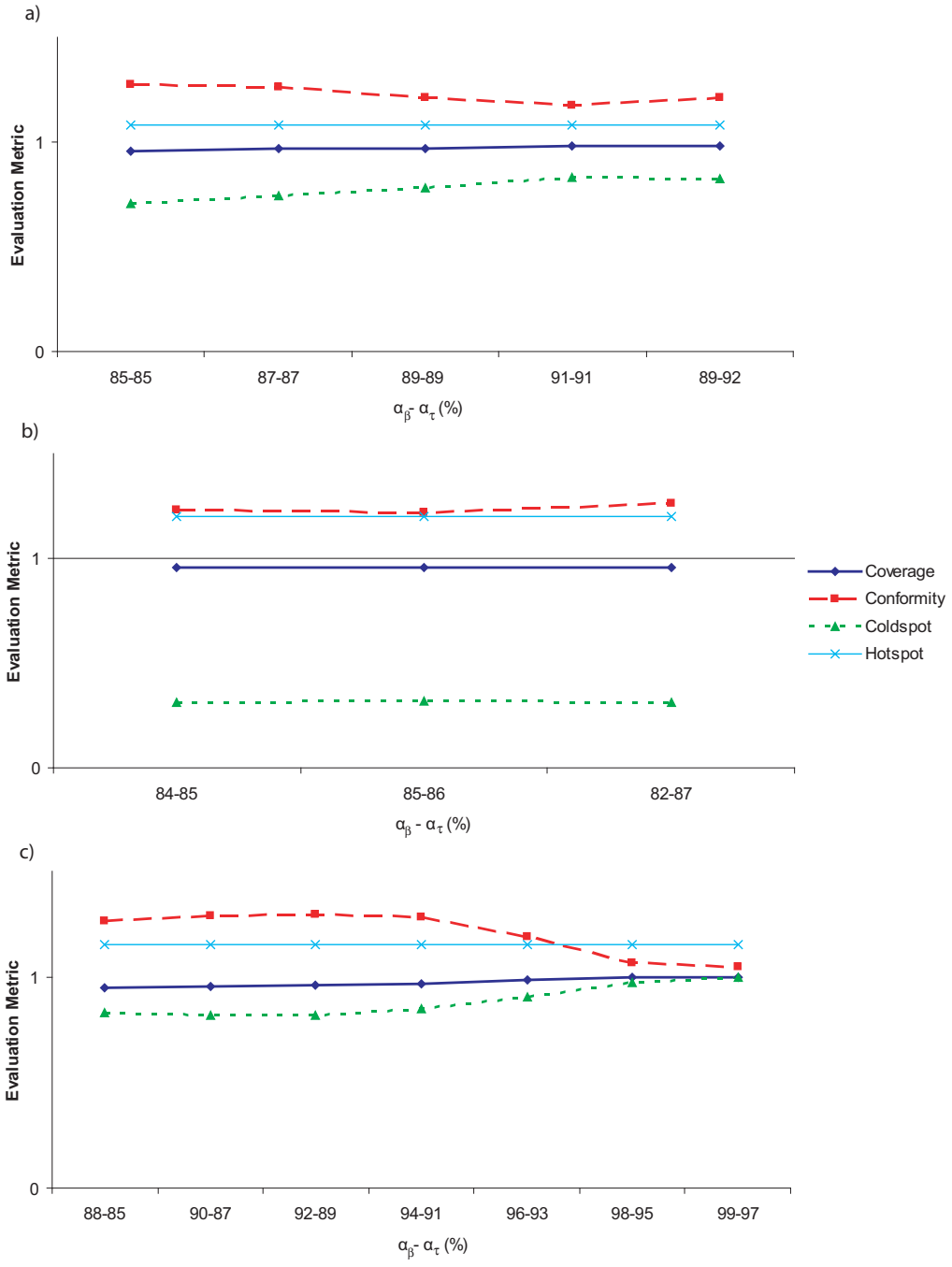

Figure 9: Change in Evaluation Metrics with Parameter Search: (a) Brain Case 8 Equi-spaced Beams, (b) Head and Neck Case 8 Equi-spaced Beams, (c) Prostate Case 6 Equi-spaced Beams

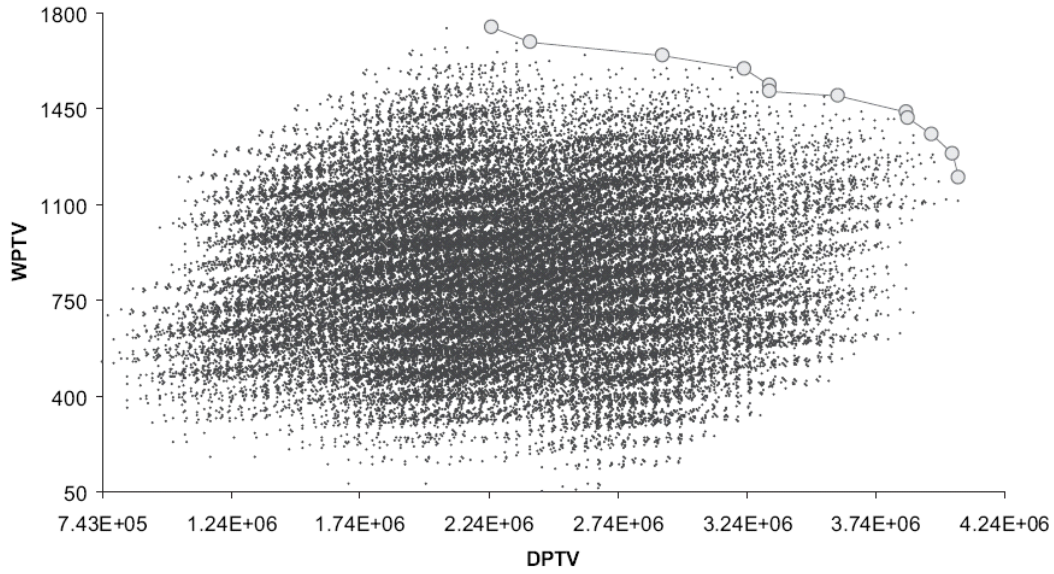

Figure 10: Brain Case - Dominated and Non-dominated Configurations

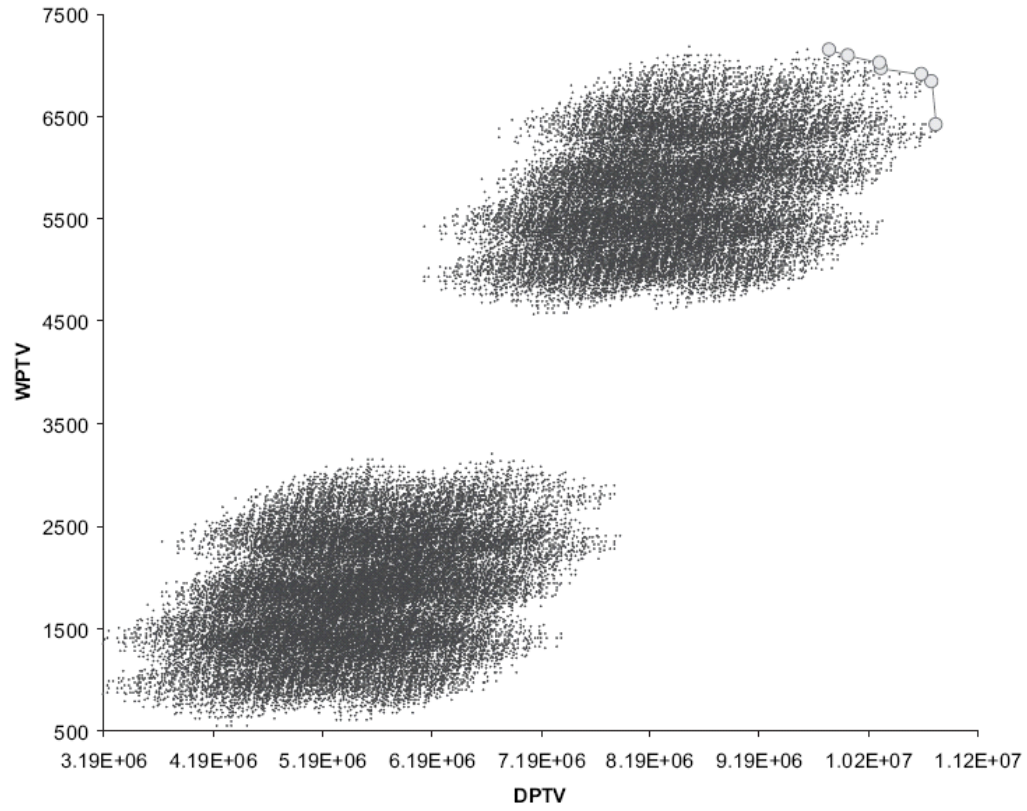

Figure 11: Head and Neck Case - Dominated and Non-dominated Configurations

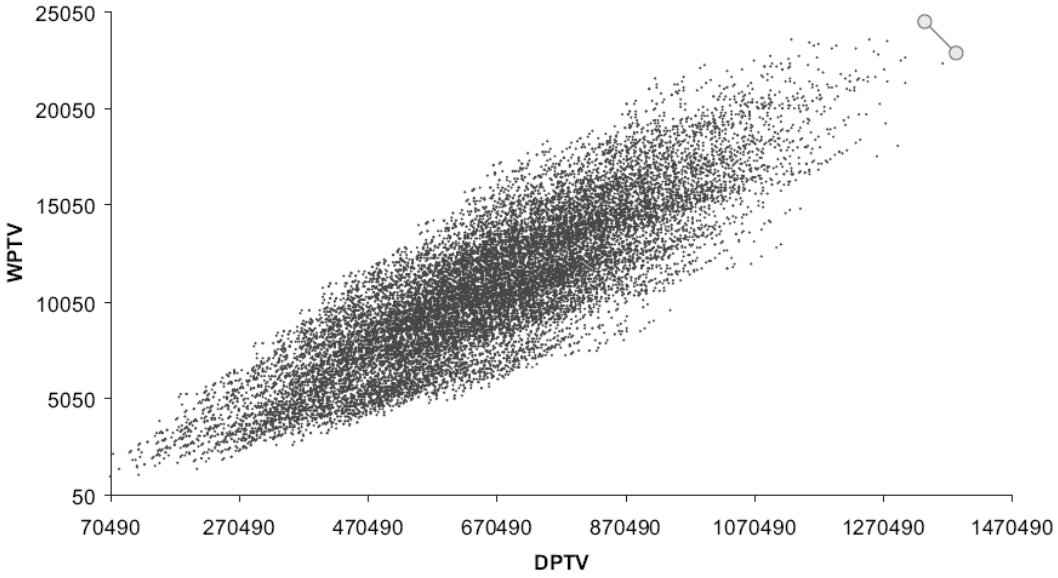

Figure 12: Prostate Case - Dominated and Non-dominated Configurations

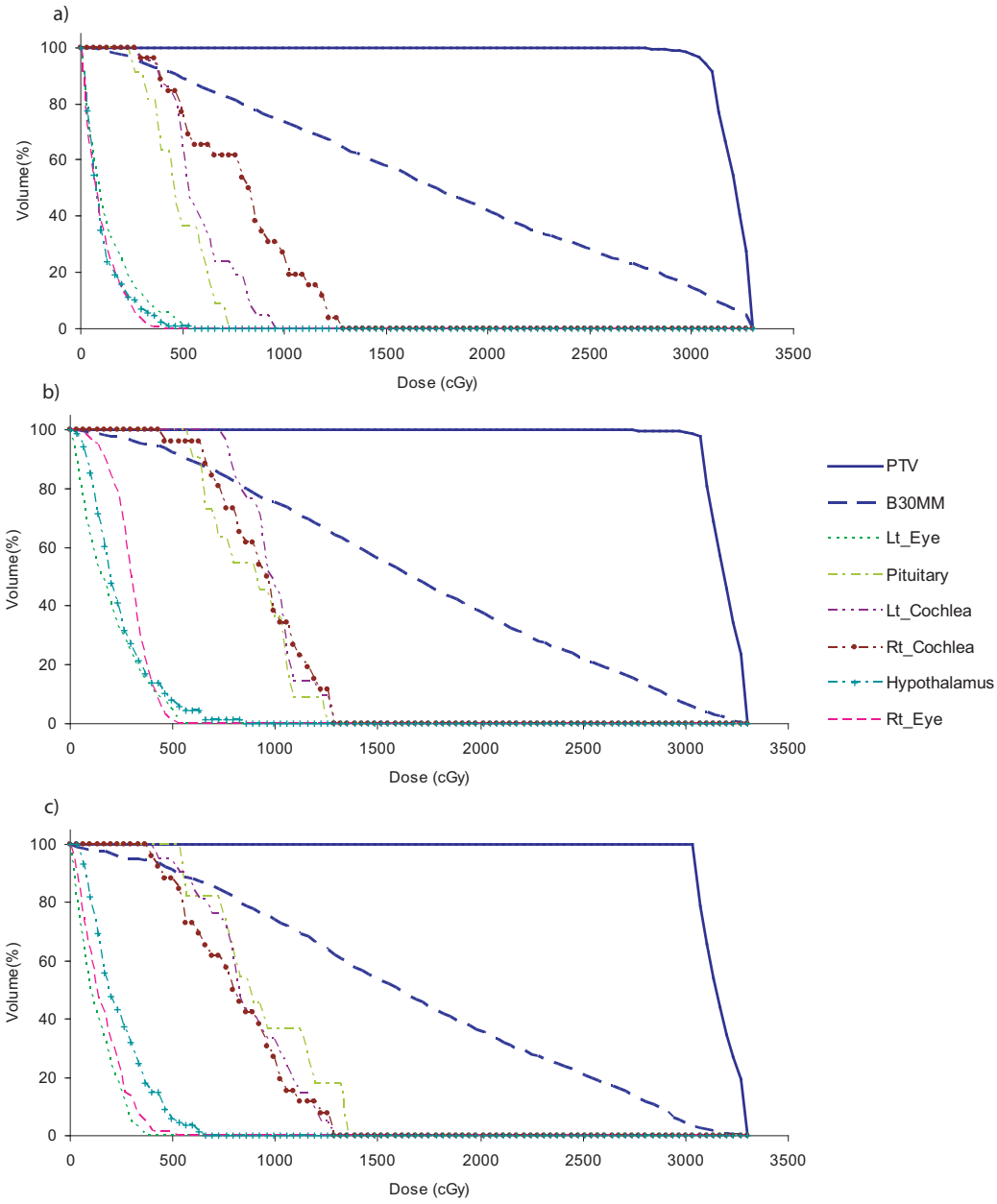

Figure 13: Brain Case Dose-Volume-Histograms: (a) Basic 8 Equi-spaced Beams Solution, (b) Final 8 Equi-Spaced Beams Solution, (c) Final 8 Selected Beams Solution

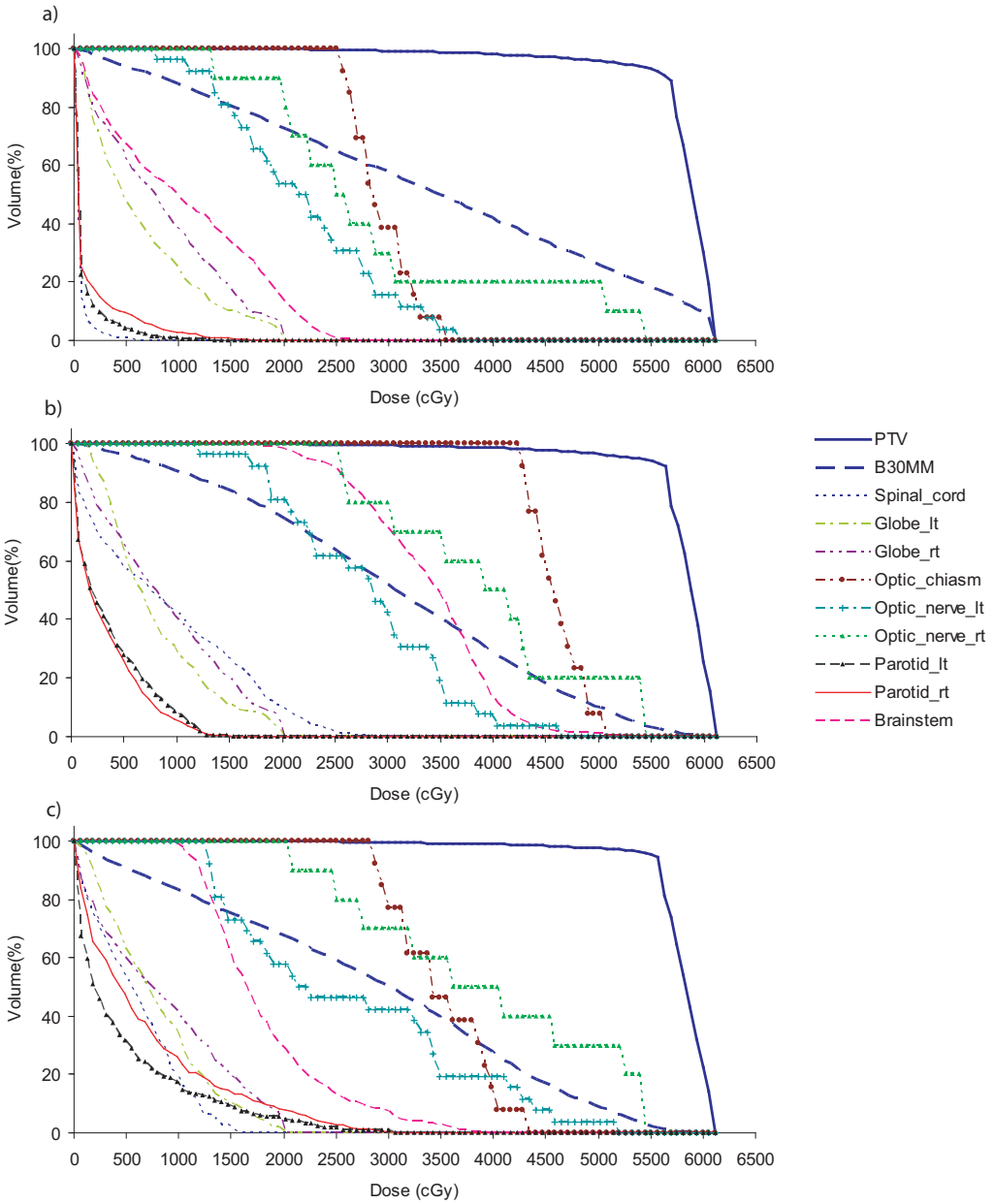

Figure 14: Head and Neck Case Dose-Volume-Histograms: (a) Basic 8 Equi-spaced Beams Solution, (b) Final 8 Equi-Spaced Beams Solution, (c) Final 8 Selected Beams Solution

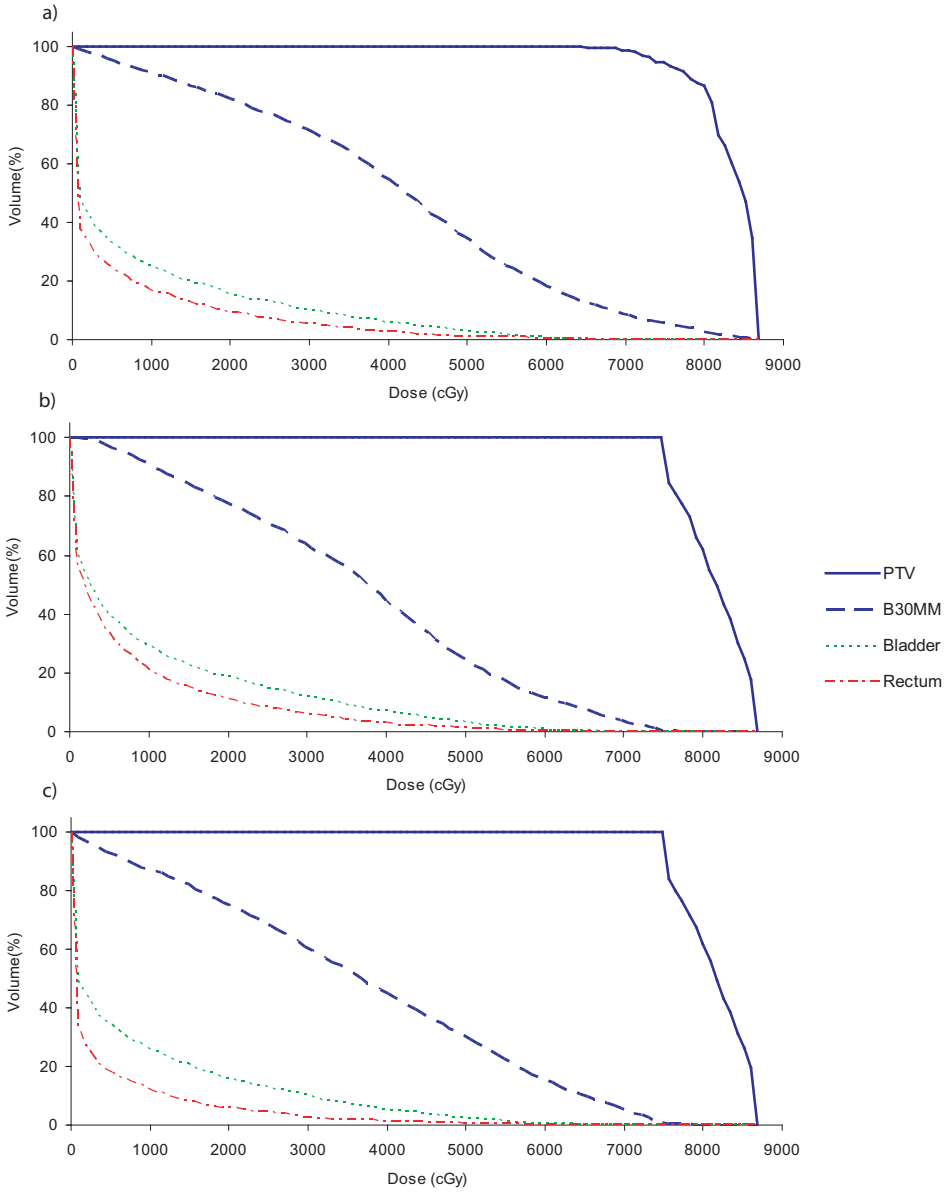

Figure 15: Prostate Case Dose-Volume-Histograms: (a) Basic 6 Equi-spaced Beams Solution, (b) Final 6 Equi-Spaced Beams Solution, (c) Final 6 Selected Beams Solution

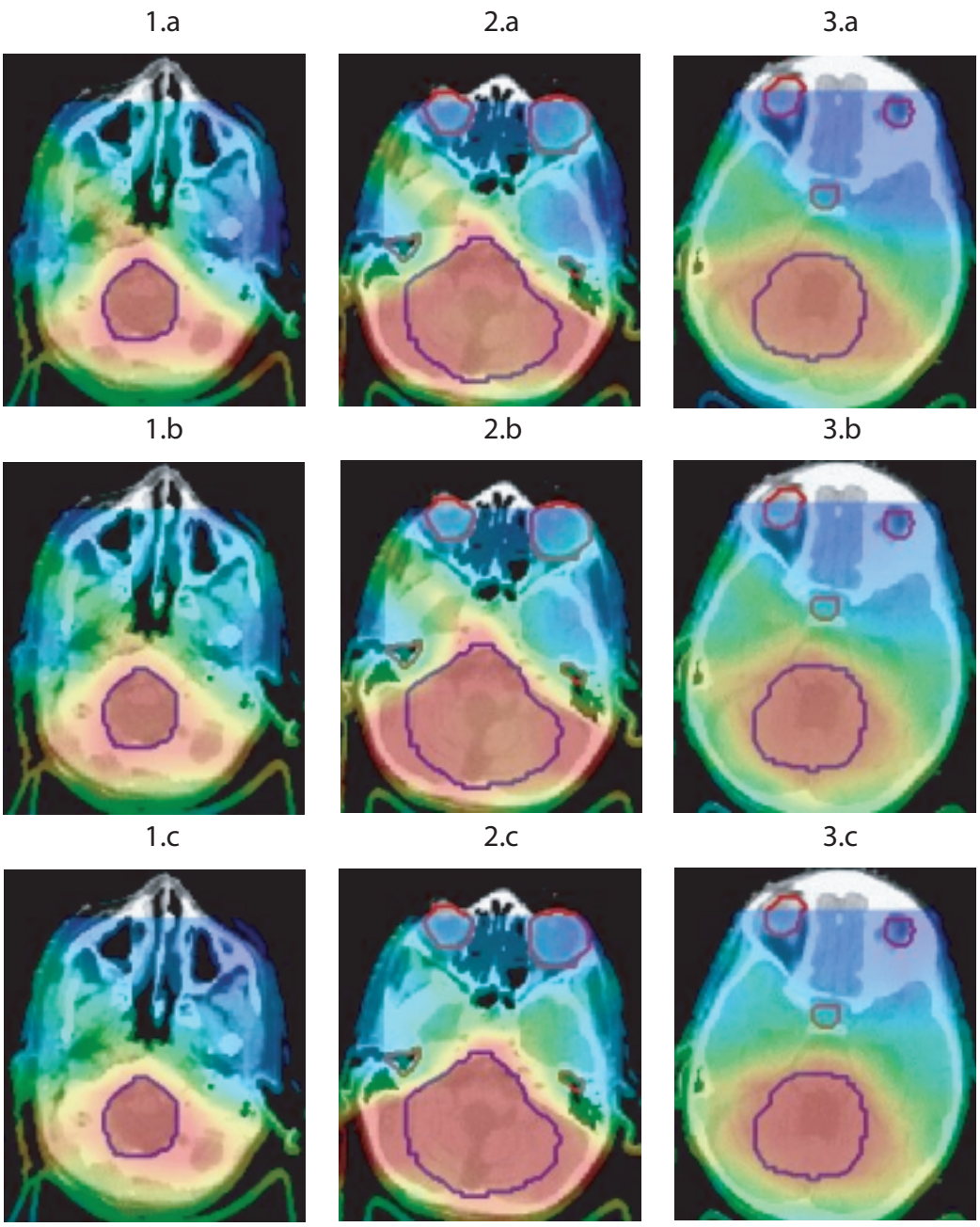

Figure 16: Brain Case Dose Distribution: (a) Basic 8 Equi-spaced Beams Solution, (b) Final 8 Equi-Spaced Beams Solution, (c) Final 8 Selected Beams Solutions are shown with a colorwash dose distribution overlaid with three CT axial images. The prescription dose is shown as red colorwash with 75, 50 and 25% of prescription dose shown as yellow, green and blue colorwash. CT axial images 1, 2, and 3 represent inferior, mid and superior slices through the target volume shown as a blue contour. The red contours represent the eyes. The cochlea structures are shown on slice 2 in red, and the pituitary is shown as red contour on slice 3.

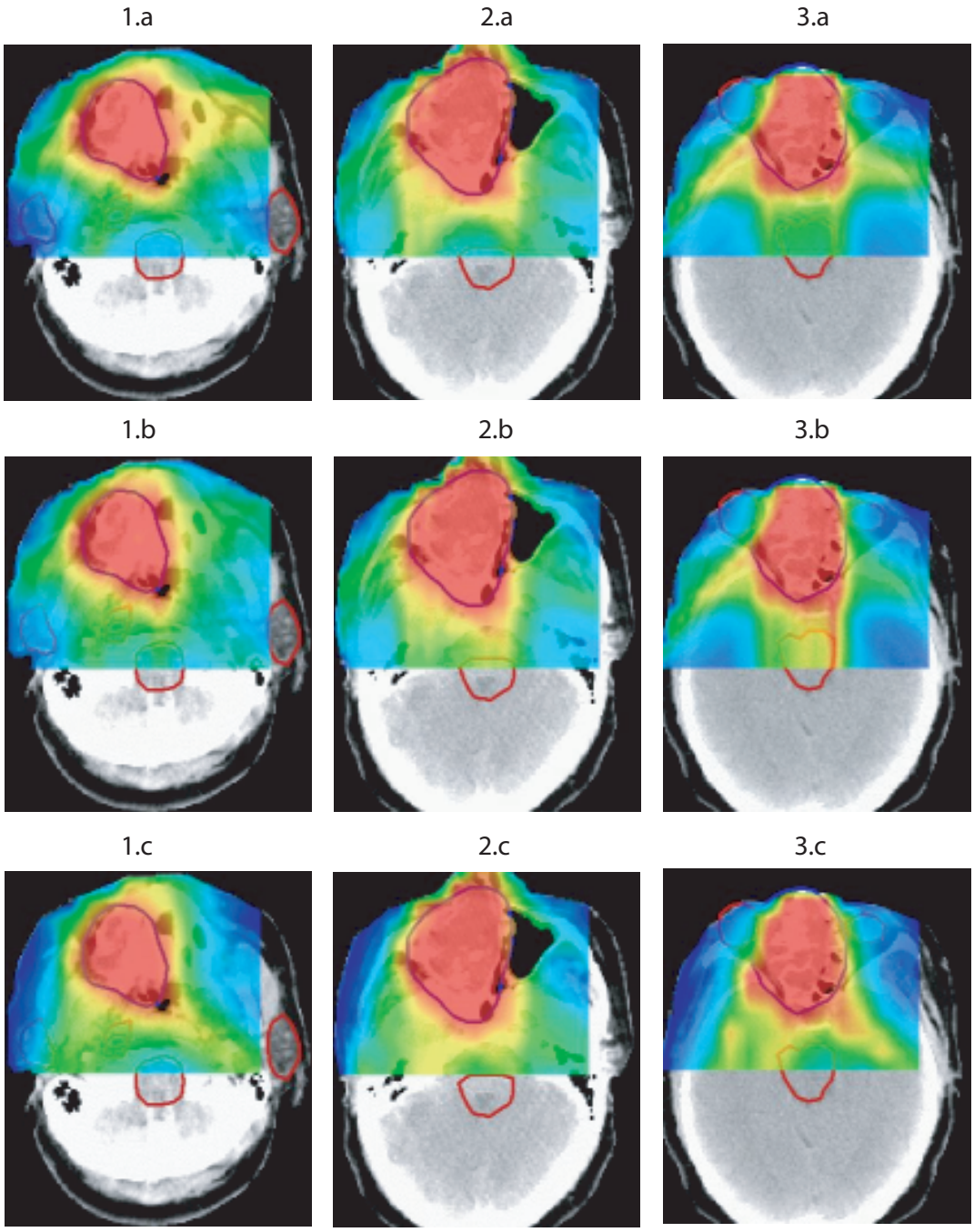

Figure 17: Head and Neck Case Dose Distribution: (a) Basic 8 Equi-spaced Beams Solution, (b) Final 8 Equi-Spaced Beams Solution, (c) Final 8 Selected Beams Solution are shown with a colorwash dose distribution overlaid with three CT axial images. The prescription dose is shown as red colorwash with 75, 50 and 25% of prescription dose shown as yellow, green and blue colorwash, respectively. CT axial images 1, 2, and 3 represent inferior, mid and superior slices through the target volume shown as a blue contour. The red central contour represents the brainstem as a critical structure. Slice 3 shows the critical structure *Globe_RT* right next to the target structure which contributes to a cold spot issue.

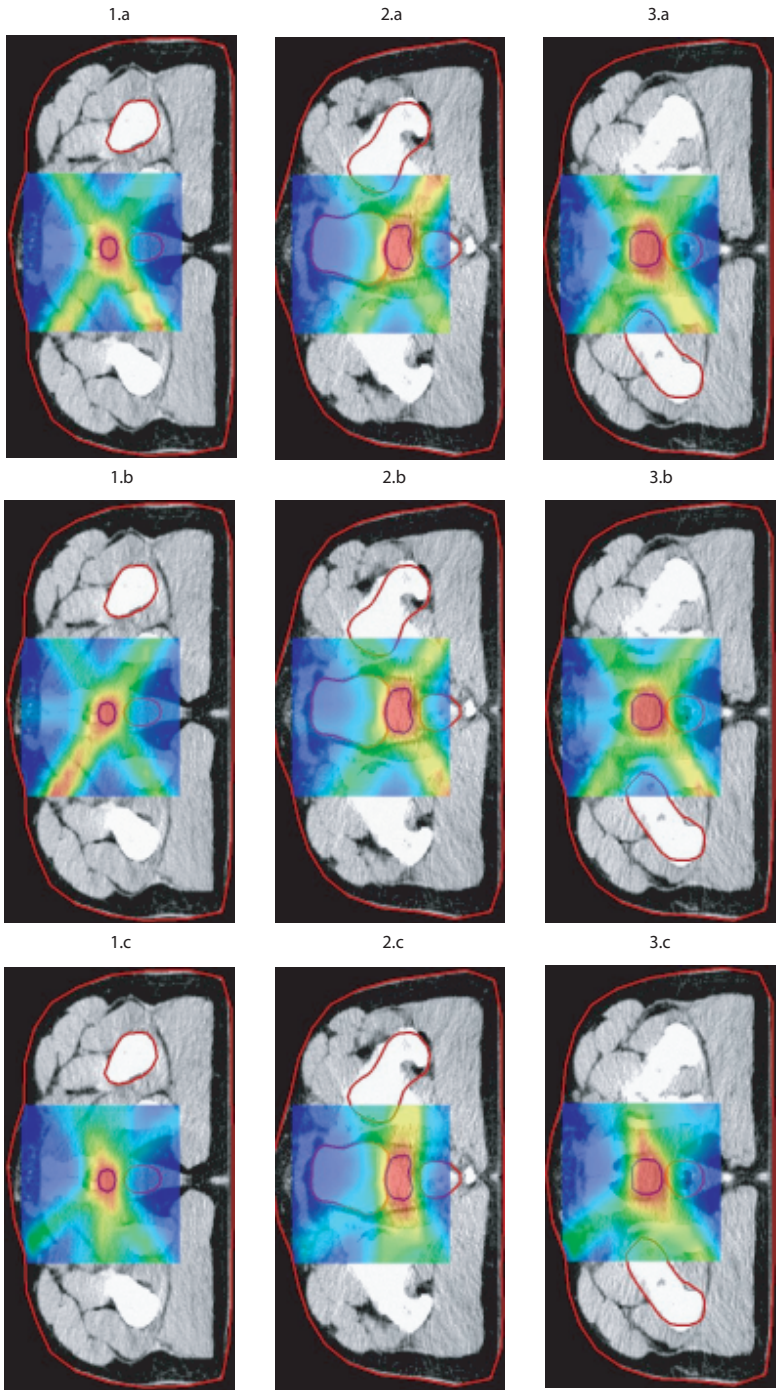

Figure 18: Prostate Case Dose Distribution: (a) Basic 6 Equi-spaced Beams Solution, (b) Final 6 Equi-Spaced Beams Solution, (c) Final 6 Selected Beams Solution are shown with a colorwash dose distribution overlaid with three CT axial images. The prescription dose is shown as red colorwash with 75, 50 and 25% of prescription dose shown as yellow, green and blue colorwash, respectively. CT axial images 1, 2, and 3 represent inferior, mid and superior slices through the prostate shown as a blue contour. The red contours on slice 2 show the bladder and rectum structures.

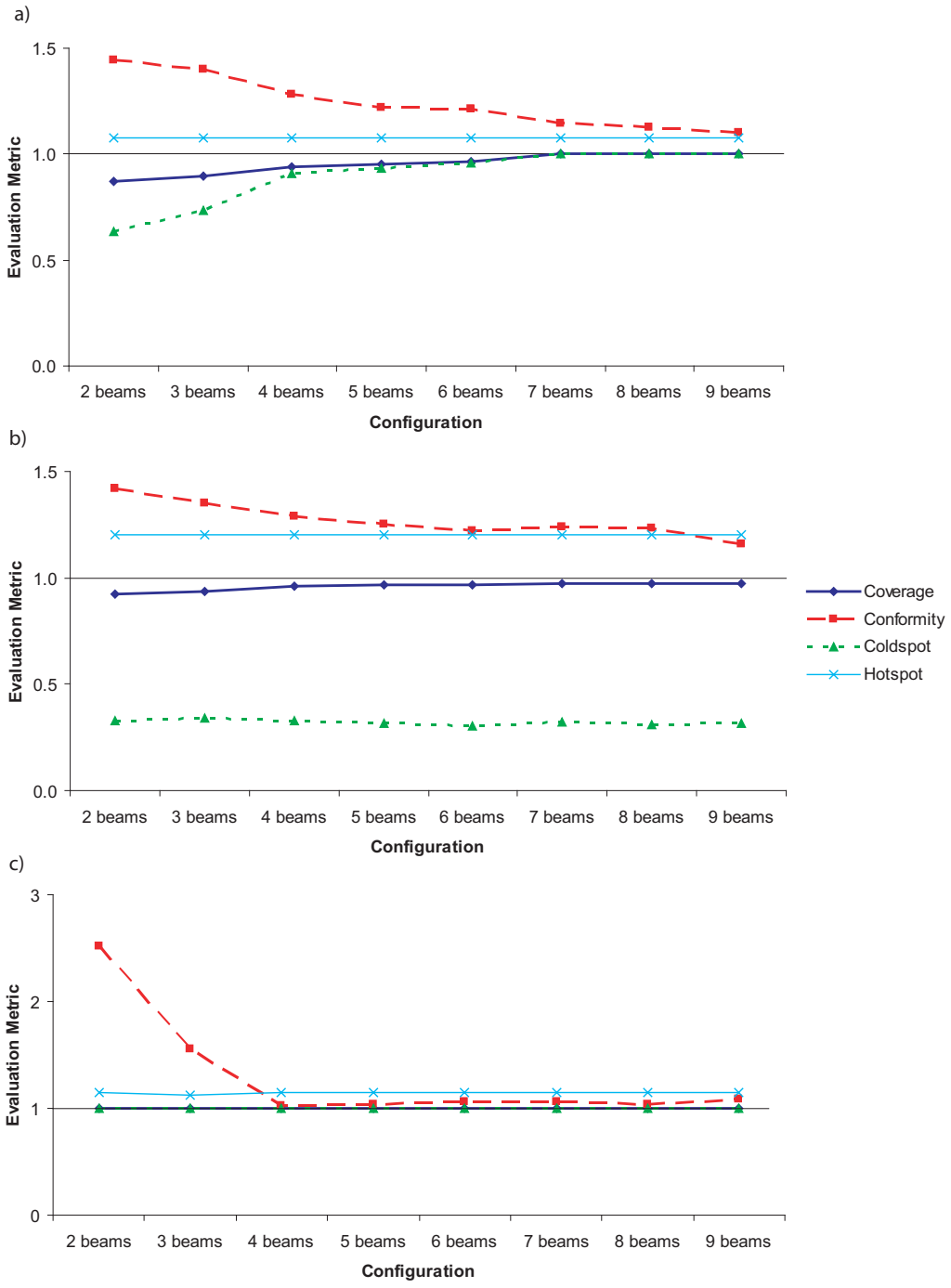

Figure 19: Final Results Returned by the Algorithm with Different Number of Beams Selected: (a) Brain Case, (b) Head andNeck, (c) Prostate

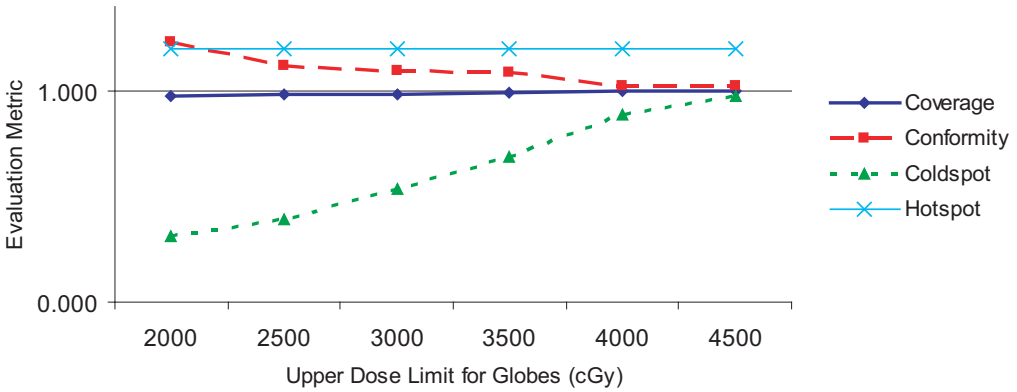

Figure 20: Head and Neck Case - Dose Limit Relaxation to Eliminate Cold Spot

CHAPTER III

MULTI-PHASE GATING IN 4D IMRT

In Chapter 2, we presented methodology to support fully automated construction of radiation treatment plans. The methodology assumes implicitly that a patient's anatomy does not change during treatment delivery. Although this assumption is reasonable for many disease sites, e.g., brain tumors, it is unwarranted for other disease sites, e.g., lung tumors. Obviously, when treating a lung tumor, the patient's anatomy changes significantly during treatment session as a result of breathing. These changes cannot be ignored and need to be considered when constructing a treatment plan. Fortunately, recent advances in imaging technology can provide multiple images of the treatment region taken at different points of the breathing cycle and deformable image registration algorithms can accurately link these images, making it possible to track an individual voxel during the entire breathing cycle. This creates the option to develop optimization models that generate a treatment plan that recognizes and distinguishes the different phases of the breathing cycle. Because of the inclusion of the time aspect, this type of treatment planning is often referred to as Four-Dimensional (4D). In this chapter, we discuss an optimization model for 4D IMRT treatment planning.

Clinical studies suggest that the organ motion during breathing is the main source of dose calculation errors in standard radiotherapy [35], as the tumor is displaced with respiratory motion with amplitudes of 5 to 12 mm in [41] which invalidates static 3D calculations on these volumes up to 9% [36] due to the interplay between target and MLC motion [54]. For example, the interplay effect [54] produces hot and cold spots [8] as the delivered dose can differ by up to 30% per field from the intended dose [17] leading to changes in the radiation's biological effects. Technological advancements allow acquiring more information about the changes in patient anatomy using the four-dimensional (4D) imaging methods [31, 45], which makes it easier to account for tumor motion. The images are typically taken

at equal time intervals during a full breathing cycle (inhale and exhale). These intervals are referred as the *phases* of the breathing cycle.

The information about organ motion is currently used at different levels of sophistication within the planning process. The simplest approach is to plan treatments on the concept of internal tumor volume (ITV), representing the envelope of the target position integrated over all phases of the breathing [25, 50]. This approach assumes unnecessarily irradiate significant amounts of normal tissue around tumor which is included to ITV and does not consider changes in dose deposition due to motion, as it assumes the target is static during delivery. Another approach is gating where the plan is constructed and delivered at only a selected phase where tumor motion is minimal [23, 47]. This results in longer treatment times as the beam is turned off most of the breathing cycle. Even with gating, selecting the delivery phase for gating is based on quantitative visual inspection rather than a real qualitative evaluation of dosimetric criteria at different phases.

Semi-empirical approaches proposed by Keall et al. [18] and Suh et al. [43], optimize the beam shapes for a single reference phase and then deform the MLC apertures according to the observed anatomical movements in other phases. Although some improvements are obtained, these approaches can track the movements only in a single dimension. Rather than optimizing for a single phase, optimization of the dose distribution in all phases is a desired goal. Lee et al. [22] propose a method that optimizes dose distribution simultaneously over all phases by creating a separate plan for each phase. Significant normal tissue sparing was observed on a simple phantom case. However, the technology to deliver multiple plans is currently not available in a clinical setting. The main advantage of this idea over gating or a series of 3D optimizations is that the unequal weighting for each phase is governed automatically by the 4D optimization algorithm and are deduced depending on the spatial relation between the target and OAR at each phase. Another obstacle towards clinical implementation of 4D-IMRT is the reproducibility of the breathing pattern observed in the images during actual delivery of the treatment.

We propose a hybrid approach where we gradually increase the number of phases to be used in treatment delivery, and simultaneously optimize the plans for the selected phases.

The algorithm picks the optimal phases to deliver the dose, with non-treatment or inactive intermediate phases in between that are used to adjust the MLC positions for the next treatment or active phase. As in Lee et al. [22], such an approach is increasing accuracy as prescription requirements are accounted over the cumulative dose in all phases. A significant challenge is the need to track voxels over multiple phases to be able to calculate the cumulative dose, repeatability of the respiratory cycle and the need of a leaf sequencing and control mechanism to deliver separate plans simultaneously.

An essential component of any 4D optimization attempt is the use of a deformable registration algorithm for tracking voxel positions and the dose received at each phase of the breathing cycle. We use an image registration technique that registers all phases of a 4D dataset simultaneously to deduce an accurate, artifact free description of the respiratory motion. The algorithm is described in detail by Schreibmann et al. [37].The use of deformable image registration to track voxels and a 4D-IMRT treatment plan optimization model has allowed us to show the tremendous potential for treatment plan quality improvement and treatment time reduction on a clinical lung case instead of artificial phantoms.

The remainder of this chapter is organized as follows. In Section 3.1 we describe multi-phase 4D treatment problem which is followed by the description of solution strategies in Section 3.2. In Section 3.3 we explore how the solution quality improves with increasing number of delivery phases with a computational study on a real lung case patient.

3.1 Problem Description

Recall that we are trying to find optimal fluence maps for each phase of the breathing cycle simultaneously by considering the overall dose delivered to patient instead of individually optimizing each phase. The deliverability of the plans is not addressed, rather we are trying to assess the benefit of using multiple phases of the respiratory cycle to improve treatment plan quality and to reduce treatment time over plans that just use single-phase gating.

To be able to calculate the total dose received by each voxel at the end of treatment it is necessary that the dose received by a voxel of a structure per intensity of a beamlet is

available for each phase and that each voxel can be tracked through the different phases. As mentioned before we use a sufficiently powerful image registration algorithms to do so. The treatment plan will specify the intensity of each beamlet in each phase and the dose received by each voxel of a structure at the end of treatment. In this setting it is possible to extend IMRT formulations described in previous chapter to optimize the total dose received by the patient over all phases.

The resulting treatment is quite different from single-phase gating as the patient is treated during multiple phases; the optimization model decides how much dose to deliver in each phase. This reduces the treatment time and thus improves the comfort level of the patient. Furthermore, because the optimization exploits the opportunities provided in each phase, better treatment plans can be obtained.

As is usual in IMRT treatment plan generation, we assume the plan requirements are in the form of a minimum prescription dose for target structure voxels and a maximum tolerance dose for nearby critical structure voxels. A prescription dose is the dose level necessary to destroy or damage target cells, while a tolerance dose is the level above which complications for healthy tissues may occur. In addition to the set of dose requirements defining acceptable treatment plans, clinical oncologists use a set of evaluation metrics to assess the quality of a treatment plan. We consider the following metrics: *coldspot*, *hotspot*, *coverage* and *conformity* as defined in Chapter 2. The goal is to come up with plans that satisfy the clinical dose requirements and have preferable values for evaluation metrics.

3.2 Methodology

We will use the following notation throughout. Let P denote the set of breathing phases, N the set of beamlets, S the set of structures, V_s the number of voxels in a structure $s \in S$, D_{ijsp} be the dose received by voxel j of structure s per unit intensity of beamlet i at phase p of the breathing cycle. Let x_{ip} be the intensity of beamlet i at phase p, i.e., a decision variable. Then the dose z_{js} received by voxel j of structure s, is

$$z_{js} = \sum_{i \in N} \sum_{p \in P} D_{ijsp} x_{ip} \qquad \forall j = 1, ..., V_s;\ s \in S. \tag{31}$$

For simplicity of exposition we assume that there is only one target structure $\tau \in S$. We extend the LP model and the parameter selection algorithm designed for 3D-IMRT planning in Chapter 2 to 4D-IMRT by replacing the formulation of z_{js} as in Equation 31 and using the extended set of decision variables x_{ip} for beamlet intensities in each phase.

The resulting LP model is,

$$\min \quad \sum_{s \in S \backslash \{\tau\}} \frac{1}{V_s} \sum_{j=1}^{V_s} z_{js} - \frac{1}{V_\tau} \sum_{j=1}^{V_\tau} z_{j\tau} \tag{32}$$

$$\text{s.t.} \quad z_{js} = \sum_{i \in N} \sum_{p \in P} D_{ijsp} x_{ip} \quad \forall j = 1, ..., V_s;\ s \in S \tag{33}$$

$$L_s \leq z_{js} \leq U_s \quad \forall j = 1, ..., V_s; s \in S \tag{34}$$

$$c_s + \frac{1}{(1-\alpha_s)V_s} \sum_{j=1}^{V_s} (z_{js} - c_s)^+ \leq U_s^\alpha \quad s \in S \backslash \{\tau\} \tag{35}$$

$$c_\tau - \frac{1}{(1-\alpha_\tau)V_\tau} \sum_{j=1}^{V_\tau} (c_\tau - z_{j\tau})^+ \geq L_\tau^\alpha \tag{36}$$

$$x_i \geq 0 \quad i = 1, ..., N \tag{37}$$

$$z_{js} \geq 0 \quad j = 1, ..., V_s; s \in S \tag{38}$$

$$c_s \text{ free} \quad s \in S. \tag{39}$$

Constraints 35 and 36 are conditional value at risk constraints that approximately model dose-volume constraints. We tune the parameters of these constraints for the target structure (PTV) and virtual critical structure (VCS) that surrounds the tumor with an automated parameter search mechanism to get plans with better coverage and conformity values. (See Chapter 2 for more details.)

3.3 Computational Study

In this section, we will apply our solution algorithm to a lung cancer case. The dose prescription is provided in Table 14. Ten CT scans of the patient have been taken with equal interval lengths during the breathing cycle. Then, the voxels are tracked through the phases with image registration and a dose matrix (D_{ijsp}) is calculated.

In order to understand the contribution of additional phases over single-phase gating, we solve the planning problem by allowing dose delivery on multiple breathing phases (e.g.

Table 14: Dose Prescriptions for Lung Case

Constraint Type	Structure	Percentage	L (cGy)	U (cGy)
Full Volume	Spinal Cord	-	0	1800
	Esophagus	-	0	2700
	Heart	-	0	3000
	Trachea	-	0	3000
	Lungs	-	0	6000
	Carina	-	0	3000
	PTV	-	0	6500
	VCS	-	0	6500
Partial Volume	PTV	95	6000	-
	Lungs	15	-	2000
	VCS	95	-	6000

deliver only in Phase 1, Phase 4 and Phase 7). The solutions provide separate fluence maps for each phase that is preselected for dose delivery and we assume that no delivery is made in other phases. Since we are not explicitly considering delivery, we introduce "buffers" between consecutive fluence maps so as to allow time to adjust MLC. A small buffer size will allow using more phases in delivery (e.g. if the buffer size is 1, the delivery in Phase 1 can be followed in Phase 3 using Phase 2 as buffer). For each buffer size all possible configurations are formed and plans are optimized by using the preselected phases in each configuration. Our goal is to show how the values of evaluation metrics improve as we increase the number of phases for delivery.

3.3.1 Buffer Size $\geq$ 5 (Single Phase Gating)

In this case only a single phase can be used for delivery. We show the values of evaluation metrics resulting from each single-phase gating solution in Table 15. The best coverage is obtained by gating in Phase 5. Note that even when considering a single phase, selecting the right phase is important as the evaluation metrics differ. The dose-volume-histogram for this solution is shown in Figure 21.

Table 15: Single-phase Gating Solutions (Buffer ≥ 5)

	P1	P2	P3	P4	**P5**	P6	P7	P8	P9	P10
Coverage	0.470	0.405	0.449	0.433	**0.516**	0.507	0.513	0.511	0.466	0.356
Conformity	1.057	1.067	1.070	1.072	**1.063**	1.061	1.060	1.063	1.067	1.075
Coldspot	0.898	0.905	0.896	0.809	**0.734**	0.835	0.875	0.513	0.860	0.841
Hotspot	1.083	1.083	1.083	1.083	**1.083**	1.083	1.083	1.083	1.083	1.084

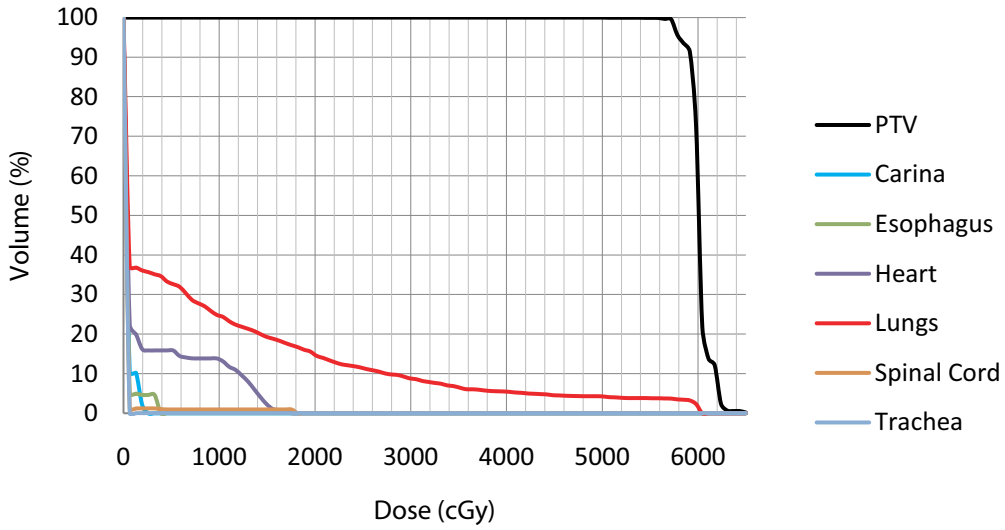

Figure 21: Single-phase Gating Solution Using Phase 5

Table 16: Two-phase Gating Solutions (Buffer = 4)

	P1-6	P2-7	P3-8	P4-9	P5-10
Coverage	**0.774**	0.751	0.733	0.637	0.725
Conformity	**1.047**	1.048	1.050	1.066	1.052
Coldspot	**0.951**	0.953	0.934	0.954	0.940
Hotspot	**1.083**	1.083	1.083	1.083	1.083

Table 17: Two-phase Gating Solutions (Buffer = 3)

	P1-5	P2-6	P3-7	P4-8	P5-9	P6-10	**P7-1**	P8-2	P9-3	P10-4
Coverage	0.746	0.744	0.755	0.695	0.645	0.741	**0.778**	0.749	0.711	0.731
Conformity	1.050	1.049	1.051	1.058	1.066	1.050	**1.047**	1.052	1.055	1.052
Coldspot	0.970	0.963	0.967	0.900	0.951	0.959	**0.944**	0.967	0.927	0.952
Hotspot	1.083	1.083	1.083	1.083	1.083	1.083	**1.083**	1.083	1.083	1.083

3.3.2 Buffer Sizes = 4 and Buffer Size = 3 (Two-Phase Gating)

For buffer sizes 4 and 3 we can deliver at most in two-phases. Tables 16 and 17 show the values of evaluation metrics resulting from two-phase gated solutions for buffer size 4 and 3, respectively. The solution with best coverage is obtained by delivering in phases 1 and 7, for which the DVH is provided in Figure 22.

3.3.3 Buffer Size = 2 (Three-Phase Gating)

If the buffer size is two, we are able to deliver using three phases of the breathing cycle. The values of evaluation metrics for each possible configuration is shown in Table 18. Using phases 1, 4 and 7 produces the best coverage. The resulting DVH for this configuration is provided in Figure 23.

Table 18: Three-Phase Gating Solutions (Buffer = 2)

	P1-4-7	P2-5-8	P3-6-9	P4-7-10	P5-8-1	P6-9-2	P7-10-3	P8-1-4	P9-2-5	P10-3-6
Coverage	**0.887**	0.846	0.839	0.871	0.846	0.871	0.868	0.863	0.862	0.867
Conformity	**1.046**	1.050	1.053	1.048	1.046	1.047	1.049	1.046	1.052	1.050
Coldspot	**0.972**	0.983	0.948	0.959	0.975	0.957	0.967	0.976	0.976	0.970
Hotspot	**1.083**	1.083	1.083	1.083	1.083	1.083	1.083	1.083	1.083	1.083

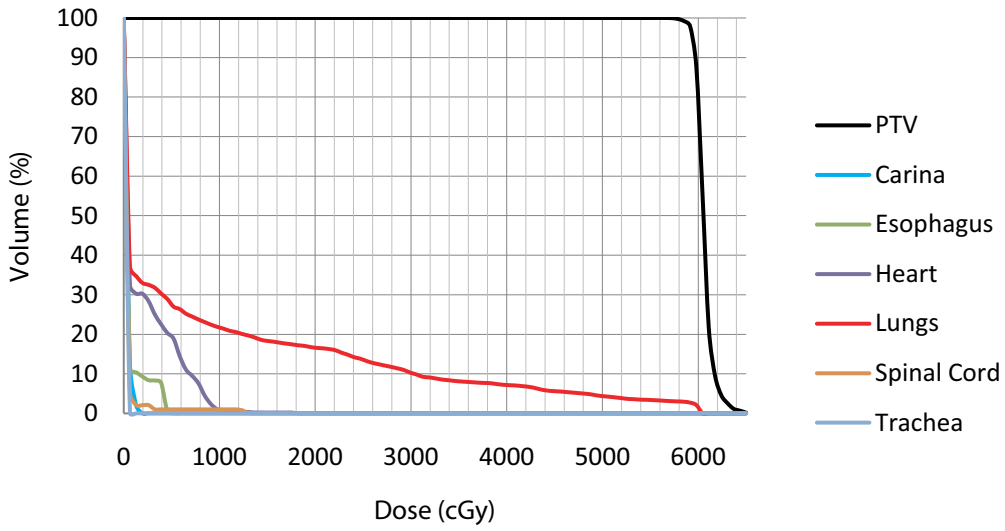

Figure 22: Two-Phase Gating Solution Using Phases 1 and 7

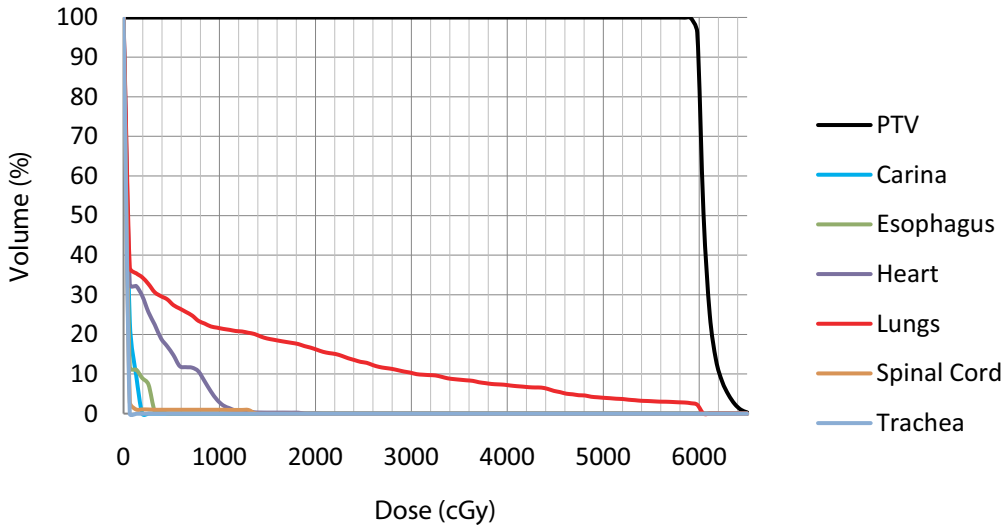

Figure 23: Three-Phase Gating Solution Using Phases 1, 4 and 7

Table 19: Five-Phase Gating Solutions (Buffer = 1)

	P1-3-5-7-9	P2-4-6-8-10
Coverage	**0.962**	0.958
Conformity	**1.046**	1.046
Coldspot	**0.992**	0.986
Hotspot	**1.083**	1.084

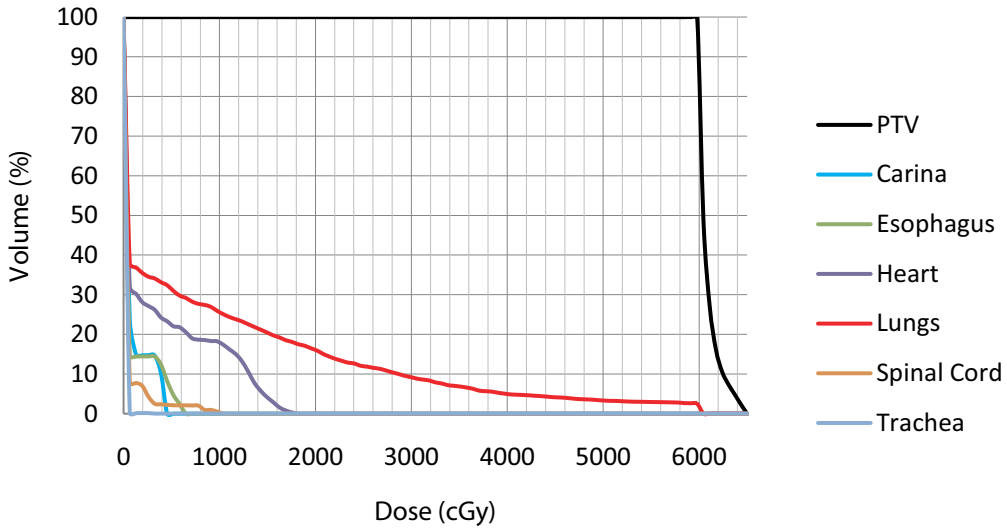

Figure 24: Five-Phase Gating Solution Using Phases 1, 3, 5, 7 and 9

3.3.4 Buffer Size = 1 (Five-Phase Gating)

In this case we can deliver at every other phase of the breathing cycle which allows delivering in five different phases. The results for the two alternatives are provided in Table 19. Delivering in phases 1, 3, 5, 7 and 9 provides a better coverage. The DVH for this configuration is shown in Figure 24.

3.3.5 Buffer Size = 0 (No-Gating)

We create a solution that allows delivery in all phases of the breathing cycle. This solution can bee thought of as the theoretical limit in decreasing the buffer size. The values of evaluation metrics are provided in Table 20 and the DVH is shown in Figure 25.

Table 20: No-Gating Solution (Buffer = 0)

	All Phases
Coverage	**0.994**
Conformity	**1.041**
Coldspot	**0.995**
Hotspot	**1.083**

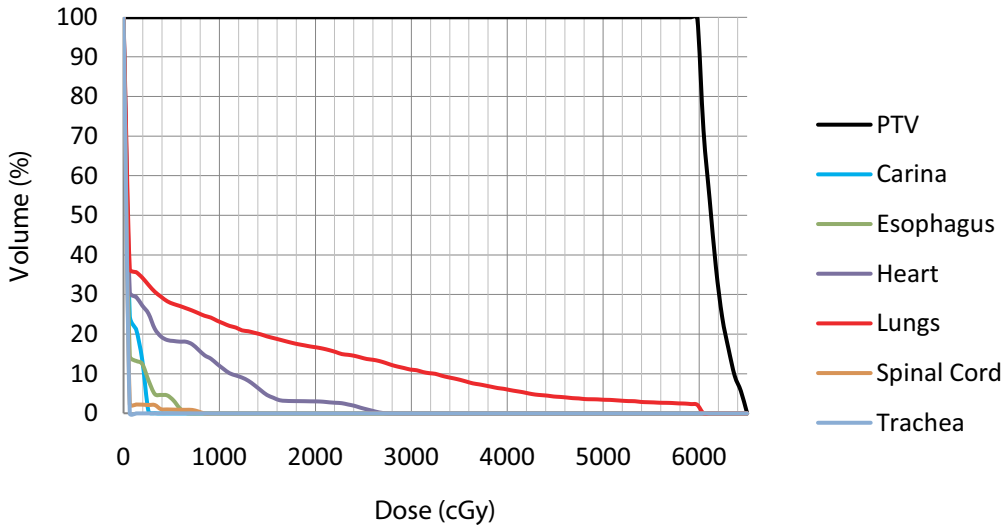

Figure 25: No-Gating Solution Using All Phases

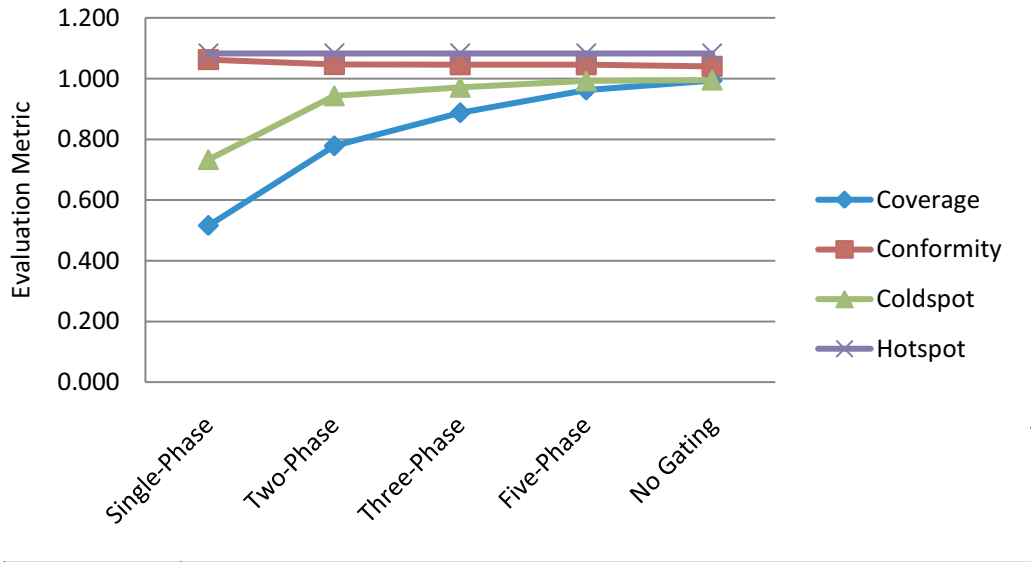

Figure 26: Change in Evaluation Metrics by Increasing Number of Delivery Phases

3.3.6 Analysis of Results

The first observation we make is that the values of the evaluation metrics improve as the number of delivery phases is increased. This is illustrated in Figure 26. The improvement is most significant for the *coverage* and *coldspot* metrics, which shows that increasing the number of phases not only increases the number of voxels that can receive the prescription dose but also increases the value of the minimum dose received by any voxel of the target structure.

The proportion of dose delivered to target structure at each phase in multi-phase solutions is not equal. The optimization algorithms decide how much dose to deliver at each phase to get the best possible dose distribution. In Figure 27, we provide the proportion of dose received by the target in a) a two-phase gating solution b) a three-phase gating solution c) a five-phase gating solution and d) no gating solution (using all phases) which are chosen as the configurations that can provide the highest coverage while obeying all dose prescription requirements.

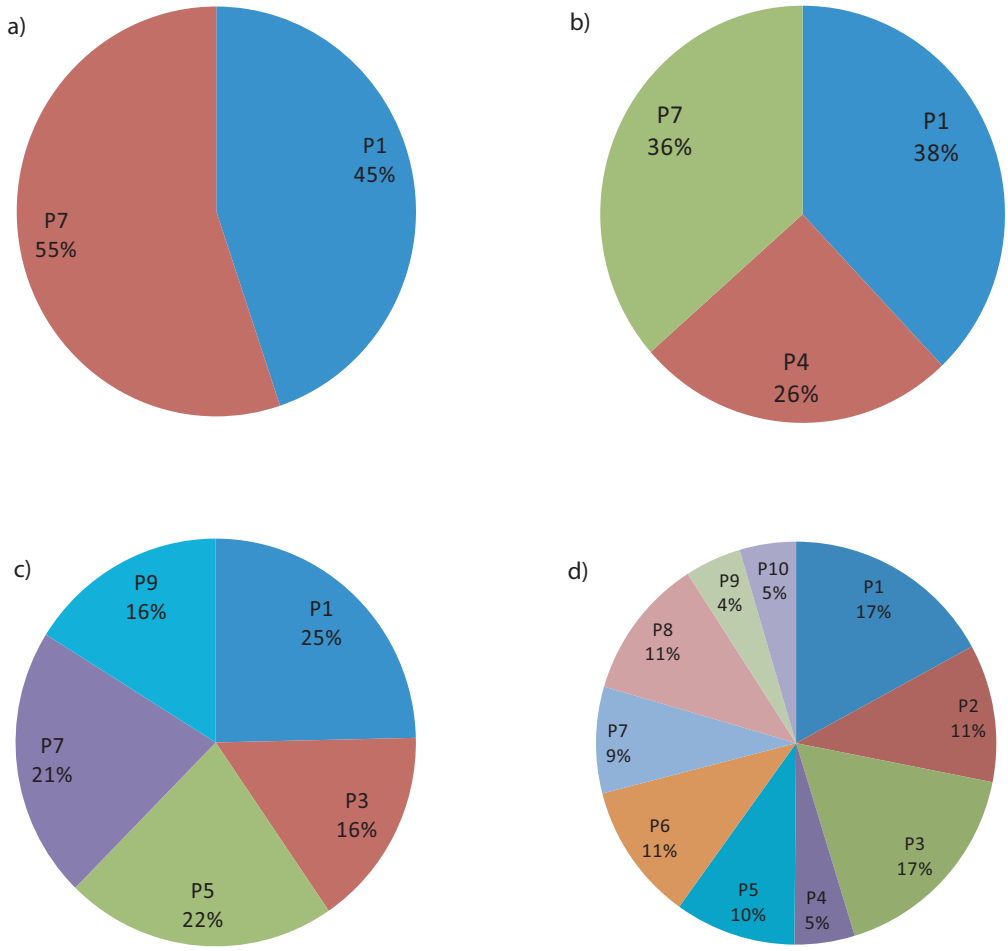

Figure 27: Proportion of Dose Received by PTV a) a two-phase gating solution b) a three-phase gating solution c) a five-phase gating solution and d) no gating solution

3.4 Concluding Remarks

In this chapter we have extended the 3D-IMRT fluence map optimization algorithms to 4D-IMRT planning using a previously developed image registration algorithm that can track an individual voxel during the entire breathing cycle. Our model finds optimal fluence maps for each phase of the breathing cycle simultaneously by considering the overall dose delivered to patient. The optimization exploits the specific opportunities provided in each of the phases and better treatment plans have been obtained. The results of this computational study shows that the tumor coverage can be improved from only 51% using single-phase gating to 96% using five-phase gating; a significant improvement in treatment effectiveness. However, the technology to deliver multiple plans is currently not available in a clinical setting. Algorithms that convert a single IMRT fluence map to a series of aperture shapes for final delivery are well developed. However, to our knowledge, no work has been done on leaf sequencing for multiple fluence maps. That is thus a natural direction for future research. The resulting technology may still produce suboptimal results due to inconsistent breathing patterns during image acquisition and actual treatment delivery. So, another direction for further research is the development of a real-time plan reoptimization technology based on the observed breathing pattern during the treatment.

CHAPTER IV

VOLUMETRIC-MODULATED ARC THERAPY TREATMENT PLANNING

Radiation treatment is delivered in daily visits lasting around five to eight weeks depending on the tumor location and size. Reducing treatment time in these sessions is important both for patient comfort and health economics as more patients can be treated. In addition, reduced treatment times decrease the chances of anatomical changes during the treatment, which lessens the discrepancy between actual and planned dose delivery. Volumetric Modulated Arc Therapy (VMAT) is a very recent radiation delivery technique that is designed to deliver high quality treatments in significantly less time than current techniques. In this chapter we describe optimization approaches for VMAT treatment planning.

Modern external-beam radiation systems contain a linear accelerator (LINAC), which can rotate around the patient as shown in Figure 1. The beams from the accelerator are shaped by a computer-controlled multi-leaf collimator (MLC) that is integrated to LINAC as shown in Figure 2. MLC contains parallel metallic leaf pairs which can be controlled individually to block some parts of the field. The open field formed by the MLC is called an *aperture* and the amount of energy delivered from the aperture per unit volume of tissue is called its *intensity*. Each aperture shape (beam shape) is modeled as a collection of pencil beams (called beamlets) and has a uniform intensity. In Chapter 2 we have presented planning approaches for Intensity-Modulated Radiation Therapy (IMRT), which produces a highly concentrated treatment of the tumor volume, while limiting the radiation dose to adjacent healthy tissue for many types of cancer (see Veldeman et al. [46]). In IMRT, the LINAC stops at several positions during its rotation to cross-fire at a cancerous tumor volume. These stopping positions for radiation delivery are called *beam angles*. At each beam angle the relative position of the patient and the machine needs to be checked for accuracy, which in practice limits the number of beam angles to be used in order to

reduce patient positioning times and chances for patient positioning errors. Thus, only a few positions of the coplanar space around the patient are exploited and at each of these positions a non-trivial amount of time is spent on activities other than radiating the tumor. To address these shortcomings of the fixed-beam IMRT approach, a rotational delivery technique called Intensity-Modulated Arc Therapy (IMAT) has been developed by Yu [53]. As opposed to IMRT where radiation is delivered only at specific beam angles, in IMAT radiation is continuously delivered with multiple overlapping arcs. IMAT is able to use a larger space around the patient, but it does not constitute a significant advantage in reducing treatment time as it requires multiple rotations around the patient. Recently, a new technique called Volumetric-Modulated Arc Therapy (VMAT) has been developed. (The technology is called VMAT by Elekta Inc. and RapidArc by Varian Medical Systems; we will use VMAT throughout this thesis.) In VMAT radiation is continuously delivered during a *single* rotation of the treatment machine around the patient. The aperture shapes and the intensity levels are continuously modulated during the rotation. The technology has been widely adopted; Elekta reports more than 200 orders in three years since its introduction. The main advantage of the VMAT technology is that the delivery is very fast, usually in less than 2 minutes compared to the treatment of the same disease site in 14 minutes with IMRT. Recent clinical studies also demonstrate VMAT's dosimetric advantage in treatment quality for cervix [10], head [14, 9], lung [3] and prostate tumors [30, 38].

For modeling purposes, as in IMRT, the structures (target and critical structures) are discretized into cubes called voxels (e.g., cubes of $5 \times 5 \times 5$ mm) and the dose delivered to each voxel per unit intensity of each beamlet (dose matrix) is required for dose calculation and planning. The total dose received by a voxel is the sum of doses deposited from each beamlet. As VMAT uses the whole coplanar space with a continuous movement around the patient, a dose matrix for each point in the rotational path is required for the most accurate dose calculations. However, this is computationally prohibitive, and therefore the dose matrix is generally calculated at *control points* for about 180 angles that are 2 degrees apart from each other. A VMAT treatment plan needs to specify a *dose rate* (intensity per time), a *rotation speed* and MLC *leaf positions* at each control point. Since there are

physical limits on the rotational speed of the linear accelerator and the speed of the MLC leaves these limits need to be considered by planning systems. A slower rotation speed provides more flexibility to create more complex shapes and a wider range for aperture intensities which may increase treatment quality while increasing treatment time.

Current implementations of the VMAT technology are based on the research of Otto et al. [29] which uses a multi-resolution approach for optimization. The work builds on previous works on IMAT optimization by different groups who have proposed algorithms to modulate the dose by the MLC collimator in arc therapy to obtain highly conformal dose distributions. Wang et al. [48] provides a review on the development of rotational techniques. Earl et al. [12] proposes a direct aperture optimization approach for IMAT that doesn't require an intermediary leaf sequencing algorithm. Cameron et al. [7] proposes a sliding window arc therapy technique and demonstrated its capability to create highly conformal treatment plans. Shepard et al. [39] proposes an algorithm that translates optimized intensity maps into deliverable IMAT plans. Ulrich et al. [44] shows that IMRT-like dose distributions can be achieved with a single arc. However, for a large number of beam angles, the planning scheme proposed by Ulrich [44] and Earl [12] can take a long time to converge and the plan quality can be limited by the constraints of MLC aperture shape connectivity. Bedford et al. [3] adapts a leaf sequencing algorithm to VMAT, which consists of a fluence optimization using the iterative least-squares technique, a segmentation and then a direct-aperture optimization originally developed for IMAT. Ma et al. [26] proposes using the scoring approach as in Pugachev and Lei [33] to provide a starting solution to the optimization algorithm.

We develop two different planning approaches for VMAT. The first approach finds leaf positions and dose rates separately in a two-stage algorithm. First, the leaf positions are selected by assigning scores to individual beamlets using their geometric and dosimetric properties, and maximizing the total score using a shortest-path algorithm in order to keep the favorable beamlets open. Once the leaf positions are selected, a linear program is solved to optimize dose rates. Although this approach can produce treatment plans very efficiently, the results are not of acceptable quality for clinical implementation. The second

approach tries to optimize leaf positions and dose rates simultaneously by solving a large scale mixed-integer programming model. As the solution time of the mixed-integer program is prohibitive, an IP-based heuristic has been developed that balances solution quality and solution time. The heuristic algorithm solves the IP model multiple times on a reduced set of of decision variables by identifying candidate leaf positions and control points at each iteration. To decrease computation times and to generate better integer solutions in a pre-specified amount of time we develop several classes of valid inequalities. Computational studies on a spinal tumor and a prostate tumor case produce clinically acceptable results.

This remainder of this chapter is organized as follows. In Section 4.1 we describe the treatment planning problem for Volumetric-Modulated Arc Therapy. Section 4.2 describes our two different solution approaches. In Section 4.3 we provide computational results from both approaches applied on a real spine and a prostate case.

4.1 Problem Description

The goal in VMAT planning is to create a dose distribution that will destroy the target cells while sparing healthy tissue by choosing optimal radiation intensities and MLC aperture shapes which can be modulated during the rotation of the linear accelerator around the patient. The dose matrix, i.e., the dose received by each voxel per unit intensity of each beamlet, is calculated at control points which are used to approximately calculate the dose delivery in a continuous rotation (Varian's RapidArc system uses 177 control points). A treatment plan needs to specify a *dose rate* (intensity per time), a *rotation speed* and MLC *leaf positions* at each control point. Machine-specific limits exist for dose rate (MU/min), gantry rotation speed (deg/sec) and MLC leaf movement speed (mm/sec). The radiation intensity between two consecutive control points depends both on the dose rate and the rotation speed. A slower rotation of the LINAC delivers a higher intensity and vice versa. The allowed change in intensity levels between two consecutive control points is a nonlinear function of the rotation speed. Since we build linear models for solution efficiency purposes, we assume a constant gantry rotation speed. Therefore, we are only interested in finding optimal intensities and leaf positions at each control point. The chosen rotation speed

guarantees that a single rotation is completed in less than two minutes. A sample solution with four control points for an MLC with two leaf pairs (blue and red) is shown in Figure 28. Each 2-dimensional array represents the beam and each cell represents a beamlet. The numbers are the intensities in monitoring units (MU). The left leaf of a pair covers the beamlets on its left side, and the right leaf covers the beamlets on its right side. At angle 0, the left leaf of blue pair covers beamlet 2 in row 1, then at angle 90 it moves one step to the left and uncovers it for delivery of 350 MUs.

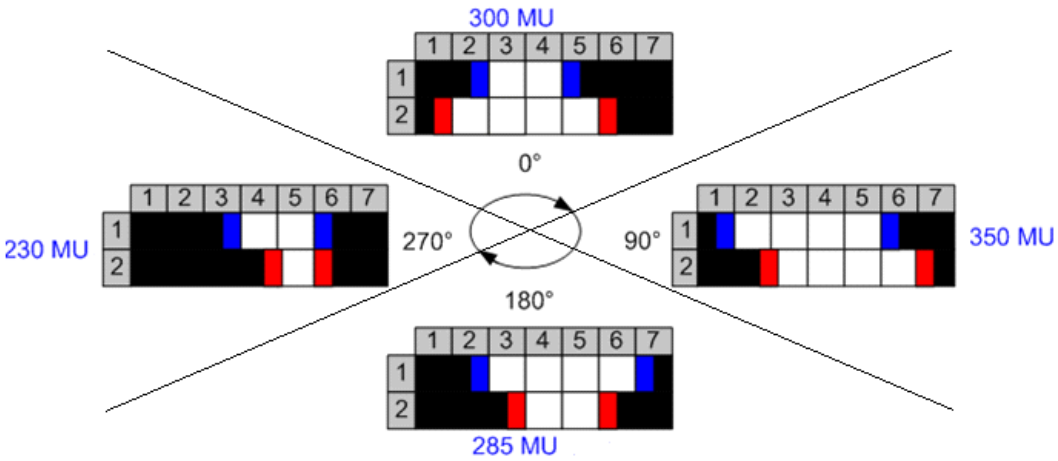

Figure 28: Sample Solution with 4 Control Points

4.2 Methodology

A plan has to specify an aperture shape formed by MLC leaves and an intensity at each control point. The first approach we describe is a two-stage algorithm in which the leaf positions and intensities are optimized separately. The second approach is based on a mixed-integer programming model where the leaf positions and intensities are optimized simultaneously.

4.2.1 Two-Stage Planning Approach

Considering the complexity of solving the planning problem for both aperture shapes and their intensities simultaneously, in our first approach, we solve the planning problem separately in two stages. In the first stage we are trying to come up with good aperture shapes at control points that satisfy technological limits related to movement of leaves

between consecutive control points. In the second stage we want to optimize radiation intensities using the aperture shapes determined in stage one.

The idea underlying the shape selection is to discern favorable beamlets, which will be used, and unfavorable beamlets (for example passing though an OAR), which will be discarded. To determine which beamlets are good, the shape selection phase starts with assigning scores to beamlets based on the open-field dose and the optimization objectives. A scoring-based beam selection algorithm by Pugachev and Lei [33] for beam angle optimization is used in this stage. After calculating individual beamlet scores, we define an aperture's score as the sum of the included beamlets' scores. Note that interaction effects are not considered in this simple summation. We also assume that leaf pairs in different rows of the collimator can move independently which is valid for most of the MLC systems. Therefore, shape selection can be solved independently for each row and these solutions can be combined to get the final shapes at each control point.

The maximum leaf movement between consecutive control points is limited due to the limited time between consecutive control points and the MLC leaf speed. Thus, the apertures should be sequenced such that the leaf movements are feasible with respect to these limits. We use a shortest path algorithm to find a feasible shape sequence that will maximize the total score. Nodes of the shortest path network are generated by enumerating possible leaf positions (assuming movements with a constant step size, e.g. 1 mm). In Figure 29, we illustrate this idea on a small example with six 1mm-width beamlets. For each node, the positions of a leaf pair and a view of the collimator are shown. These sets of nodes are generated at every control point.

Then, the sets of nodes between consecutive control points are connected according to the feasible connections shown in the rightmost columns of Figure 29. Note that in this example the leaf movement between consecutive control points is assumed to be at most 1 mm and that the leaves can either move 1 mm or stay at the same position. If there is a feasible movement option from a node in control point i to a node in control point $i+1$ then there is an arc connecting node i to node $i+1$. The selection of an arc (i, j) in the optimal solution of shortest path problem corresponds to the selection of the shape

			Beamlets						
Node	Left Leaf	Right Leaf	1	2	3	4	5	6	Feasible Connections
1	0	0							1 2 8
2	0	1	■						1 2 3 8 9
3	0	2	■	■					2 3 4 8 9 10
4	0	3	■	■	■				3 4 5 9 10 11
5	0	4	■	■	■	■			4 5 6 10 11 12
6	0	5	■	■	■	■	■		5 6 7 11 12 13
7	0	6	■	■	■	■	■	■	6 7 12 13
8	1	1							1 2 3 8 9 14
9	1	2		■					2 3 4 8 9 10 14 15
10	1	3		■	■				3 4 5 9 10 11 14 15 16
11	1	4		■	■	■			4 5 6 10 11 12 15 16 17
12	1	5		■	■	■	■		5 6 7 11 12 13 16 17 18
13	1	6		■	■	■	■	■	6 7 12 13 17 18
14	2	2							8 9 10 14 15 19
15	2	3			■				9 10 11 14 15 16 19 20
16	2	4			■	■			10 11 12 15 16 17 19 20 21
17	2	5			■	■	■		11 12 13 16 17 18 20 21 22
18	2	6			■	■	■	■	12 13 17 18 21 22
19	3	3							14 15 16 19 20 23
20	3	4				■			15 16 17 19 20 21 23 24
21	3	5				■	■		16 17 18 20 21 22 23 24 25
22	3	6				■	■	■	17 18 21 22 24 25
23	4	4							19 20 21 23 24 26
24	4	5					■		20 21 22 23 24 25 26 27
25	4	6					■	■	21 22 24 25 26 27
26	5	5							23 24 25 26 27 28
27	5	6						■	24 25 26 27 28
28	6	6							26 27 28

Figure 29: Network Nodes and Arcs

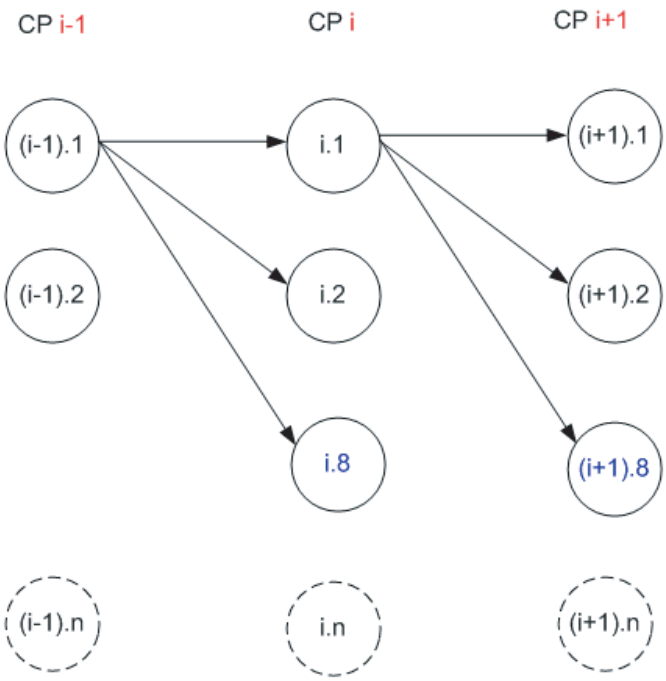

Figure 30: Network Nodes and Arcs

at node j following the shape at node i. Therefore, the arc cost between nodes i and j is the contribution of node j to the solution which is a function of the score of aperture shape j. The resulting graph can be topologically ordered and shortest path algorithms can be solved efficiently in $O(m)$ time, m being number of arcs. In Figure 30, there is a partial view of the network for the example described showing connections between control points $i-1$, i and $i+1$, where there are n nodes generated at each control point.

After finding shape sequences the second stage is to determine radiation intensities. Given the shapes, this problem is similar to IMRT planning, where given a set of shapes the goal is to decide the dose rates. The aperture intensities are optimized using the parameter selection scheme described in Chapter 2 to improve coverage and conformity while minimizing average dose delivered to critical structures.

The algorithm has been tested on a real prostate case. Although this method is very efficient, the plan quality is not good enough. Even in the case when most of the dose-related constraints for the critical structures are relaxed a significant portion of the target structure is still underdosed as shown in the dose-volume-histogram in Figure 31. This is due to the

fact that the aperture shapes and their intensities are not optimized simultaneously, and the interaction effects among multiple beamlets are ignored in determining their scores and resulting aperture shapes.

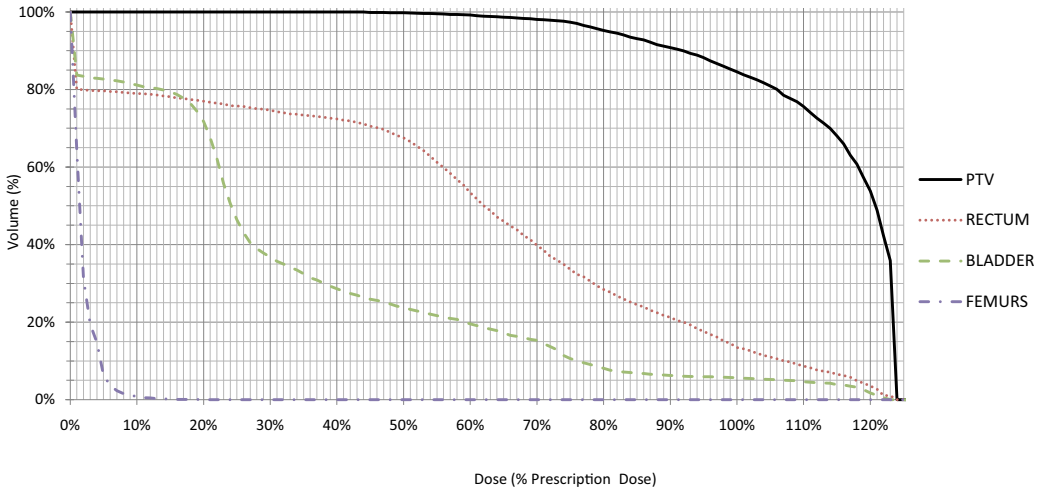

Figure 31: Dose-Volume Histogram for a Prostate Case Solution with 2-Stage VMAT Algorithm

4.2.2 Single Stage (MIP-Based) Approach

The two-stage algorithm fails to produce solutions of acceptable quality because it ignores interaction between multiple beamlets. Therefore, in this section, we describe a method based on a mixed integer programming model that is able to evaluate the interaction effects and determine aperture shapes and intensities simultaneously. We will use the following notation. Let I denote the number of control points, R the number of MLC rows (leaf pairs), C the number of MLC columns (MLC opening is discretized to columns for modeling), S the set of structures, V_s the number of voxels in a structure $s \in S$. Let D_{irc}^{js} be the dose received by voxel j of structure s per unit intensity of the beamlet in row r and column c at control point i. For simplicity of exposition we assume that there is only one target structure $\tau \in S$. Indices i, r and c are illustrated in Figure 32. The decision variables and the constraints of the problem are defined next.

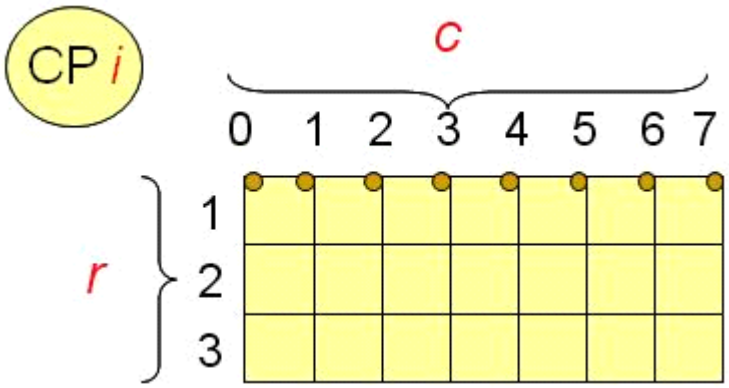

Figure 32: Control Points and Indices

Decision variables:

$$x_i \quad : \quad \text{intensity at control point } i$$

$$z_{js} \quad : \quad \text{dose received by voxel } j \text{ of structure } s$$

$$y_{irc}^{left} = \begin{cases} 1 & \text{if the left leaf at control point } i \text{ and row } r \text{ is at position } c, \\ 0 & \text{otherwise} \end{cases}$$

$$y_{irc}^{right} = \begin{cases} 1 & \text{if the right leaf at control point } i \text{ and row } r \text{ is at position } c, \\ 0 & \text{otherwise} \end{cases}$$

$$w_{irc} = \begin{cases} x_i & \text{if beamlet at control point } i, \text{ row } r, \text{ and column } c \text{ is open}, \\ 0 & \text{otherwise} \end{cases}$$

Dose related full volume and approximate partial-volume constraints using C-VaR are implemented, as before, for the plans to satisfy the prescription requirements. Let L_s and U_s be the prescribed lower and upper dose limits for structure s, respectively. The full-volume constraints for dose limits are:

$$L_s \leq z_{js} \leq U_s \qquad \forall j = 1, ..., V_s; \ s \in S. \tag{40}$$

Partial-volume constraints specify dose limits that have to be satisfied by a specified fraction of the voxels of a structure. For target structure τ these constraints are formulated as the

C-VaR constraint:

$$c_\tau - \frac{1}{(1-\alpha_\tau).V_\tau} \sum_{j=1}^{V_\tau} (c_\tau - z_{j\tau})^+ \geq L_\tau^\alpha \tag{41}$$

where $(a)^+$ is equal to a if $a \geq 0$ and 0 otherwise. This C-VaR constraint enforces that the average dose in the $(1-\alpha_\tau)$-fraction of voxels of target structure τ receiving the *lowest* amount of dose is *greater* than or equal to L_τ^α. If satisfied, at least $\alpha_\tau \times 100$ per cent of the voxels receive a dose greater than or equal to L_τ^α. Here c_τ is a free variable associated with the dose-volume constraint of structure τ. Similarly, the partial-volume constraints for critical structures are formulated as:

$$c_s + \frac{1}{(1-\alpha_s).V_s} \sum_{j=1}^{V_s} (z_{js} - c_s)^+ \leq U_s^\alpha \qquad s \in S. \tag{42}$$

For critical structure s the average dose in the subset of size $(1-\alpha_s)$ receiving the *highest* amount of dose is required to be *less* than or equal to U_s^α. When satisfied at least α_s percent of the voxels will receive a dose less than or equal to U_s^α.

Technological constraints for leaf movements are modeled as follows assuming a maximum change of δ columns between two consecutive control points. δ is calculated prior to optimization using gantry rotation speed (deg/sec), leaf movement speed (mm/sec), spacing between control points (deg), and width of MLC columns (mm):

$$\sum_{c=0}^{C} y_{irc}^{left} = 1 \quad i = 1, ..., I;\ r = 1, ..., R \tag{43}$$

$$\sum_{c=0}^{C} y_{irc}^{right} = 1 \quad i = 1, ..., I;\ r = 1, ..., R \tag{44}$$

$$y_{irc}^{left} \leq \sum_{p=MAX(0,\,c-\delta)}^{MIN(c+\delta,\,C)} y_{(i+1)rp}^{left} \quad i = 1, ..., I-1;\ r = 1, ..., R;\ c = 0, ..., C \tag{45}$$

$$y_{irc}^{right} \leq \sum_{p=MAX(0,\,c-\delta)}^{MIN(c+\delta,\,C)} y_{(i+1)rp}^{right} \quad i = 1, ..., I-1;\ r = 1, ..., R;\ c = 0, ..., C \tag{46}$$

Constraints (43) and (44) ensure that a position is chosen for each leaf pair (left and right). Constraints (45) and (46) force that the change is in leaf position between two consecutive control points is within the technological limits for left and right leaves, respectively.

A beamlet indexed by $\{i, r, c\}$ delivers dose with intensity w_{irc} if it is not covered by any of the leaves in the MLC. The dose received by voxels, z_{js}, is calculated using the following

sets of constraints:

$$w_{irc} \leq M \sum_{p=0}^{c-1} y_{irp}^{left} \quad i = 1, ..., I;\ r = 1, ..., R;\ c = 1, ..., C \tag{47}$$

$$w_{irc} \leq M \sum_{p=c}^{C} y_{irp}^{right} \quad i = 1, ..., I;\ r = 1, ..., R;\ c = 1, ..., C \tag{48}$$

$$w_{irc} \leq x_i \quad i = 1, ..., I;\ r = 1, ..., R;\ c = 1, ..., C \tag{49}$$

$$w_{irc} \geq x_i + M(\sum_{p=0}^{c-1} y_{irp}^{left} + \sum_{p=c}^{C} y_{irp}^{right} - 2) \quad i = 1, ..., I;\ r = 1, ..., R;\ c = 1, ..., C \tag{50}$$

$$z_{js} = \lambda \sum_{i=1}^{I} \sum_{r=1}^{R} \sum_{c=1}^{C} w_{irc} D_{irc}^{js} \quad s = 1, ..., S;\ j = 1, ..., V_s. \tag{51}$$

The scalar λ converts dose intensity (MU) to cGy. The constant M is a large enough number to model logical constraints and can be chosen to be equal to the maximum dose rate for x_i. Constraints (47) and (48) ensure that a beamlet doesn't deliver any radiation if it is covered by the left leaf or the right leaf and has zero intensity. Constraints (49) and (50) force the intensities of the open beamlets at the same control point to be equal to each other. Constraint (51) calculates the total dose delivered to a voxel from all beamlets.

The radiation intensity from each control point is limited by the minimum and maximum dose rates which depends on the design of the linear accelerator (e.g. minimum rate = 0.1 MU/min and maximum rate = 600 MU/min). Of course, the intensity also depends on the rotation speed (e.g. 240 deg/min) and the spacing between control points (which determines how much time is spent when traveling between two consecutive control points). Let $MinInt$ and $MaxInt$ be the minimum and maximum intensity levels that can be delivered from each control point. Then

$$MinInt \leq x_i \leq MaxInt \quad i = 1, ..., I. \tag{52}$$

Next, we provide some valid inequalities for the MIP model described above. These valid inequalities eliminate certain solutions to the linear programming relaxation of the MIP and may therefore decrease solution times:

$$\sum_{c=1}^{C} w_{irc} \leq MaxInt(\sum_{c=0}^{C} (c) y_{irc}^{right} - \sum_{c=0}^{C} (c) y_{irc}^{left}) \quad i = 1, ..., I;\ r = 1, ..., R. \tag{53}$$

Inequality (53) is valid, because $\sum_{c=0}^{C} (c) y_{irc}^{left}$ and $\sum_{c=0}^{C} (c) y_{irc}^{right}$ correspond to the position of

left and right leaves, respectively in an integer solution. Then the difference $\sum_{c=0}^{C}(c)y_{irc}^{right} - \sum_{c=0}^{C}(c)y_{irc}^{left}$ is the number of beamlets open in that row of the collimator. The inequality limits the total value of beamlet intensities in that row using the maximum intensity value, $MaxInt$, and number of open beamlets.

The following inequality is also valid:

$$w_{ir1} + \sum_{c=1}^{C-1} |w_{ir(c+1)} - w_{irc}| + w_{irC} \leq 2x_i \quad i = 1, ..., I;\ r = 1, ..., R. \tag{54}$$

The terms on the left hand-side calculate the sum of changes in beamlet aperture values from left-most beamlet to the right-most beamlet in a row. In an integer solution this should be less than twice the intensity at that control point, because the intensity level can only increase and decrease once (increase to x_i right next to the left leaf position and decrease to zero after the right leaf position).

On a clinical spine case with a 20x100 collimator and 177 control points; this model has 715,080 binary variables and more than two million constraints and cannot be solved in clinically allowable time limits. In the following section, we develop an iterative heuristic solution algorithm based on this MIP model.

4.2.2.1 MIP-Based Heuristic Algorithm

The MIP model contains all the decision alternatives for both leaf positions and dose rates, and incorporates all technological and dose-related constraints. However, the solution time required for this model is prohibitive for a clinical application, and therefore we develop a heuristic algorithm that balances solution quality and solution time. The algorithm is iterative and considers only a limited set of alternative leaf positions at each iteration, which decreases the size of the model and enables us to get integer solutions in an acceptable amount of computation time.

The number of binary variables increases with the number of control points, number of MLC rows, and number of MLC columns ($2IR(C+1)$). Our idea is based on the fact that by focusing only on a limited set of alternative decisions we can get problem instances that can be solved efficiently. The limited set should be chosen carefully without oversimplifying the

capabilities of the treatment system. First, we focus on the number of control points where the plans need to specify the delivery parameters. The current implementation of VMAT in Varian's clinical system requires 177 control points separated by about 2 degrees for a good approximation of the continuous delivery. However, this doesn't necessarily mean that the radiation treatment must be delivered from all control points. As discussed in Chapter 2, there are diminishing returns when the number of beam angles are increased. Therefore, we focus on identifying promising beam angles and eliminate some of the control points in order to have a solvable instance of the model while balancing solution quality.

We use beams-eye-view (BEV), i.e. a 2-dimensional view of the treatment area as seen from the linear accelerator, at each control point to set the initial aperture shapes. The leaf positions are chosen to create the smallest opening that contains the target area as shown in Figure 33. Let the initial positions of the left and right leaves of a leaf pair r at control point i be BEV_{ir}^{left} and BEV_{ir}^{right}, respectively. When the MLC leaf positions are fixed the model

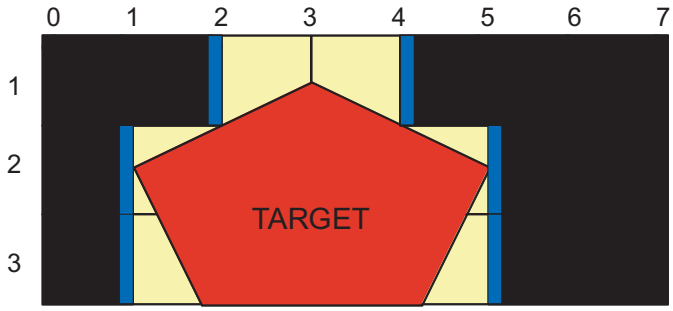

Figure 33: Using Beams-eye-view for Initial Aperture Shapes

reduces to a linear programming model with aperture intensities (x_i) being the remaining decision variables. We optimize this model with an objective function that maximizes the minimum dose on the target structure while satisfying all dose related constraints. Then, we compare the optimized aperture intensity values at each control point. The idea is to discern control points around which (by also looking at neighboring control points) the optimal solution delivers only a negligible (very close to minimum value, $MinInt$) amount of radiation. If any control point and its immediate neighbors have only negligible intensities, that control point is not included in the optimization model in the following iteration. Of

course, this initial selection of control points assumes a very large MLC opening without any leaf modulation other than closing the areas outside target beams-eye-view. We address this issue by updating the control point selection at the end of each iteration.

Elimination of some control points decreases the model size almost in half in the real-life test cases that we have worked on (e.g. 177 to 85 in a prostate case), but still the MIP model is too large. Therefore, we develop mechanisms to focus on a reduced set of candidate leaf positions at each iteration to further decrease problem size. First of all, at each row, we do not consider the leaf positions which are outside the BEV of the target structure in order to prevent delivery from the beamlets which do not pass through the target. Then, the candidate leaf positions for the right and left leaf of leaf pair r at control point i have to be chosen from the set:

$$BEV_{ir} = \{c : BEV_{ir}^{left} \leq c \leq BEV_{ir}^{right}\} \tag{55}$$

For example, in Figure 33 only positions 2, 3 and 4 can be picked as candidates for the leaf pair at row 1.

We define the following parameters to guide the candidate leaf position selection from set BEV_{ir} at each iteration:

Neighborhood: The width (measured by number of columns) of the search region to pick candidate leaf positions, which is denoted by N^t at iteration t.

Step-size: The number of columns between two consecutive candidate leaf positions, which is denoted by E^t at iteration t.

Let LT_{ir}^t and RT_{ir}^t be the positions of the left and right leaf for the leaf pair r at control point i at the *beginning* of iteration t, respectively. Then, the sets of candidate positions for the left and right leaves at iteration t are defined as:

$$ALT(Left)_{ir}^t = \{c : c = MAX(LT_{ir}^t - kE^t, BEV_{ir}^{left}),\ 0 \leq k \leq \frac{N^t}{E^t},\ k \in \mathbb{Z}\} \cup$$
$$\{c : c = MIN(LT_{ir}^t + kE^t, BEV_{ir}^{right}),\ 0 \leq k \leq \frac{N^t}{E^t},\ k \in \mathbb{Z}\}$$

and

$$ALT(Right)^t_{ir} = \{c : c = MAX(RT^t_{ir} - kE^t, BEV^{left}_{ir}),\ 0 \leq k \leq \frac{N^t}{E^t},\ k \in \mathbb{Z}\} \cup$$
$$\{c : c = MIN(RT^t_{ir} + kE^t, BEV^{right}_{ir}),\ 0 \leq k \leq \frac{N^t}{E^t},\ k \in \mathbb{Z}\}, \text{ respectively.}$$

$MAX(a, b)$ equals to the maximum of numbers a and b, and $MIN(a, b)$ equals to the minimum of numbers a and b. The candidate columns are selected in the neighborhood of the previous solution by scanning the area with the specified step-size. If the neighborhood expands beyond the target beams-eye-view then the bounding columns, i.e. BEV^{left}_{ir} and BEV^{right}_{ir}, are added to the candidate sets. Using this scheme the number of binary variables required to specify the decision alternatives for a leaf pair r at iteration t will be at most:

$$2\frac{N^t}{E^t} + 1.$$

Thus, either the neighborhood, N^t, should be smaller or the step-size, E^t should be higher to have a smaller number of binary variables. To have a similar number of binary variables at each iteration we choose neighborhood and step-size parameters such that the ratio $\frac{N^t}{E^t}$ is constant throughout the iterations.

We start with a high value of N^t to scan a larger area in the earlier iterations and decrease it gradually. In Figure 34 we illustrate how the resolution and neighborhood for candidate leaf position selection works for three consecutive iterations until a final solution is found. The size of the neighborhood is decreased gradually and chosen as twice the value of step-size at each iteration. In Iteration 1, we start with the leaves chosen at BEV positions as described above. The candidate leaf positions for the left leaf and right leaf are shown with blue and red circles, respectively. The starting positions of the leaves are shown with rectangles. Note that the candidates are selected based on these starting positions, neighborhood and step-size.

In the earlier iterations, due to low resolution, we cannot expect to get high quality solutions that satisfy all the dose-related prescriptions. Therefore, we introduce a mechanism that relaxes the dose-related constraints and tightens them gradually as the iterations progress. Let the relaxation factor for critical structures be CR^t and the relaxation factor

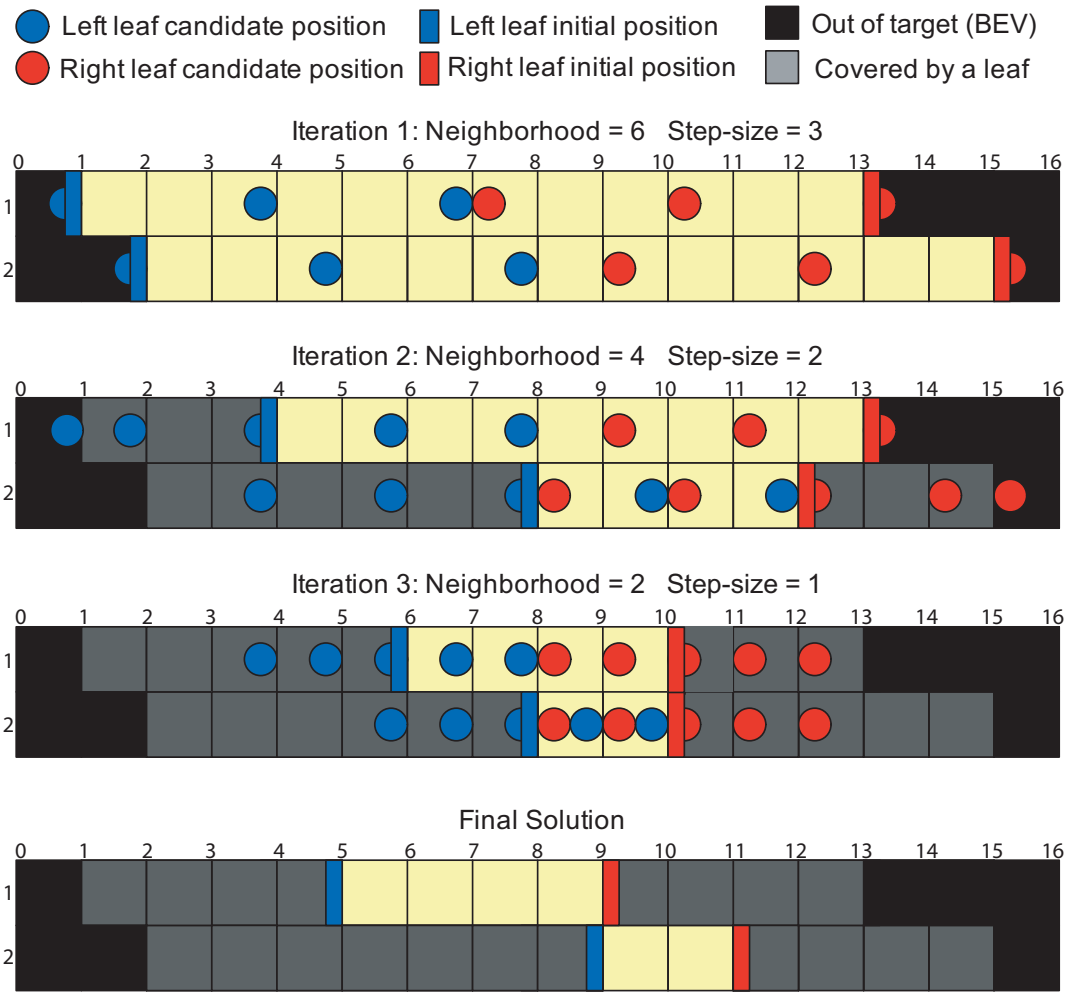

Figure 34: Selection of Candidate Leaf Positions

for target structures be TR^t at iteration t. Then the upper dose limit, U_s, for a critical structure s is increased to $(1 + CR^t)U_s$, and the lower dose limit for a target structure τ is decreased to $(1-TR^t)L_\tau$. The relaxation parameters for critical structures are selected such that CR^t converge to zero to force the actual dose constraints in the final iteration. Ideally, the value of TR^t should also converge to zero to satisfy target radiation limits, however a value less than 5% is generally acceptable by the clinicians (e.g. 95% of the prescription dose). In summary, we control the algorithm by 4 parameters (N^t, E^t, CR^t and TR^t) which are specified at each iteration to guide the algorithm to pick the decision alternatives and to relax the prescription requirements as necessary.

We can integrate various objective functions formed by costlets that measure plan quality

(see Kessler et al. [19]). In this algorithm, considering that the target radiation is forced by dose-volume constraints, we minimize only the weighted sum of average dose on critical structures:

$$\sum_{s \in S \setminus \{\tau\}} \frac{1}{V_s} \sum_{j=1}^{V_s} z_{js}. \tag{56}$$

When the step-size is greater than 1, the distance between two consecutive candidate positions may be greater than one. In this case, we make the following observation. The beamlets which are surrounded by the same candidate leaf positions will have equal intensities (denoted by w_{irc}) in an integer solution. Let two consecutive candidate leaf positions be denoted by the pair (k, l). We introduce the following set of valid inequalities:

$$w_{irc} = w_{ir(c+1)} \qquad k < c < l;\ \forall (k,l) \text{ pairs};\ i = 1, ..., I;\ r = 1, ..., R. \tag{57}$$

We have introduced three sets of valid inequalities in Equations (53), (54) and (57). In Tables 21 and 22 we computationally test the performance of all possible combinations of the set of valid inequalities in the first iteration of the algorithm. The performance measure is defined as the value of objective function after a run time of 15, 30, 45 and 60 minutes. We observe that using Equation (53) by itself does not contribute to improved performance, although it slightly improves the value of the lower bound. The configuration which contains Equations (54) and (57) produces the best results. In the computational studies, which we present in the next section, this configuration of valid inequalities is used in all iterations of the algorithm.

4.3 Computational Results

In this section, we present the results of a computational study of two clinical cases: a spine case and a prostate case. The computational study focuses on the value of the MIP-based heuristic introduced above. The cases are intentionally picked from different disease sites to test the robustness of the algorithm. The prescriptions for the case are provided in Tables 23 and 24. The modulated area of the MLC is 24×60mm for the spine case and 20X100mm for the prostate case, and the number of control points is 177. We used the MIP

solver of XPRESS and run the experiments on a 2.66GHz Pentium Core 2 Duo processor with 3GB of RAM under Windows XP Operating System.

The algorithm starts by setting the leaf positions to target structure beams-eye-view. Then the iterations follow as described in the methodology. The values of the parameters that we selected for the neighborhood size (N^t), step-size (E^t), critical structure limits relaxation factor (CR^t), and target structure relaxation factor (TR^t) at each iteration can be found in Table 25. We start with a large step size of 5 columns and iteratively decrease it to 1 to do the final iterations on a grid with higher resolution to get a more precise modulation of the leaves. The target tolerance factor could only be reduced to 4% which forces 95% of the target structure to receive 96% (1-4%) of the prescription dose. The selection of these parameters has been done manually and a better selection mechanism for these parameters may be the subject of future research. For example, the tolerance factors should be balanced for different cases; higher values of the tolerance factors for critical structures may avoid infeasibility at earlier iterations, but may lead to infeasibility in the later iterations. The tolerance factors may also be customized for each structure individually by using the geometric and dosimetric information.

At each iteration the solution time is limited to 900 seconds, which bounds the total solution time to 135 minutes for 9 iterations. In fact, in some iterations the integer solutions may be found quicker and the algorithm switches to next iteration earlier. The iterations are also terminated if an integer solution with a gap of 2% is found. Tables 26 and 27 provide the solution times for spine and prostate cases, respectively.

The dose-volume-histogram for the final solution of the spine case is shown in Figure 35. The curves are acceptable for both the critical structures and the target structure indicating that the algorithm is successful in creating a VMAT plan in a planning time of approximately two hours. In Figures 36-39 we present the change in the curves over nine iterations. We observe that the target (PTV) DVH moves upwards meaning that a better target radiation is obtained gradually as intended by the gradual tolerance variation.

Similarly, the dose-volume-histogram for the final solution of the prostate case is shown in Figure 40. Again, the curves are acceptable for both the critical structures and the target

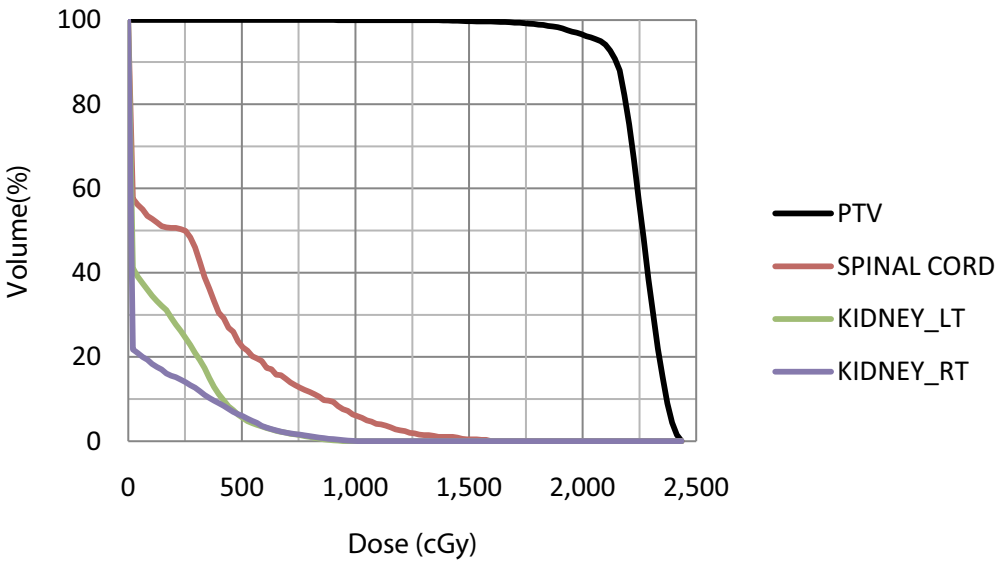

Figure 35: The Final Dose-Volume-Histogram for the Spine Case

structure indicating that the algorithm is successful in planning for both cases. In Figures 41-44 we present the change in the curves over nine iterations. We observe that the target (PTV) DVH moves upwards meaning that a better target radiation is obtained gradually as intended by the gradual tolerance variation.

4.4 Concluding Remarks

In this chapter we have built a large-scale mixed-integer programming model for VMAT treatment plan optimization. Because the solution of this model was computationally prohibitive we have developed an iterative MIP-based heuristic algorithm which solves the model multiple times on a reduced set of decision variables. We introduced valid inequalities that decrease solution times, and, more importantly, that identify higher quality integer solutions within specified time limits. Computational studies on a spinal tumor and a prostate tumor case has produced clinically acceptable results. Future research should focus on even more efficient solution of the resulting integer programs. Various branching and node selection options can be explored or other strong valid inequalities may be introduced. Selection of decision variables may be improved by considering other geometrical and dosimetric information, e.g., by classifying rows and control points and developing

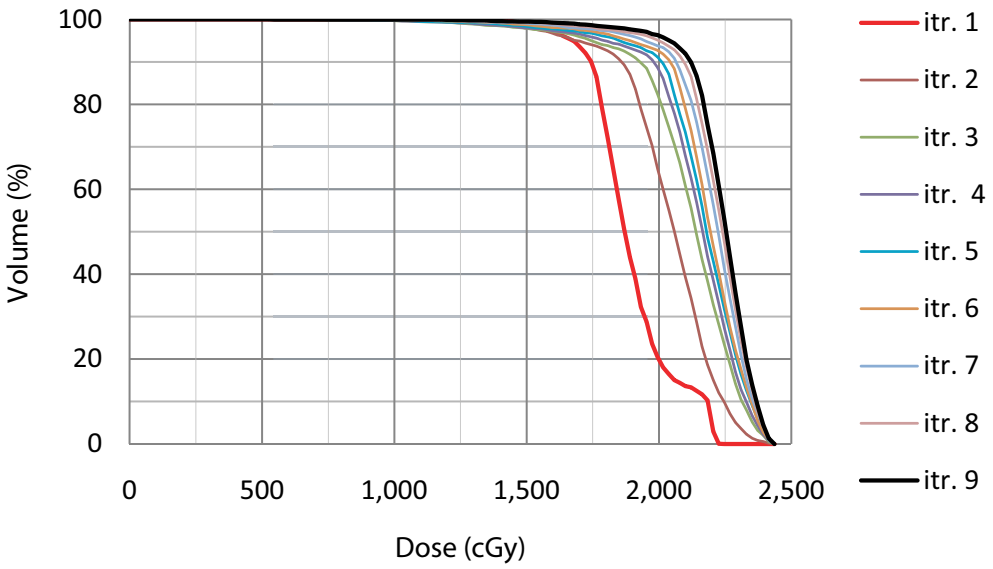

Figure 36: Change in PTV (Spine Case) Dose-Volume-Histogram over the Iterations

more customized candidate position selection mechanisms. The trade-off between solution efficiency and the approximation errors in dose calculation may be explored to come up with a balanced number of MLC rows, columns and voxel sample sizes.

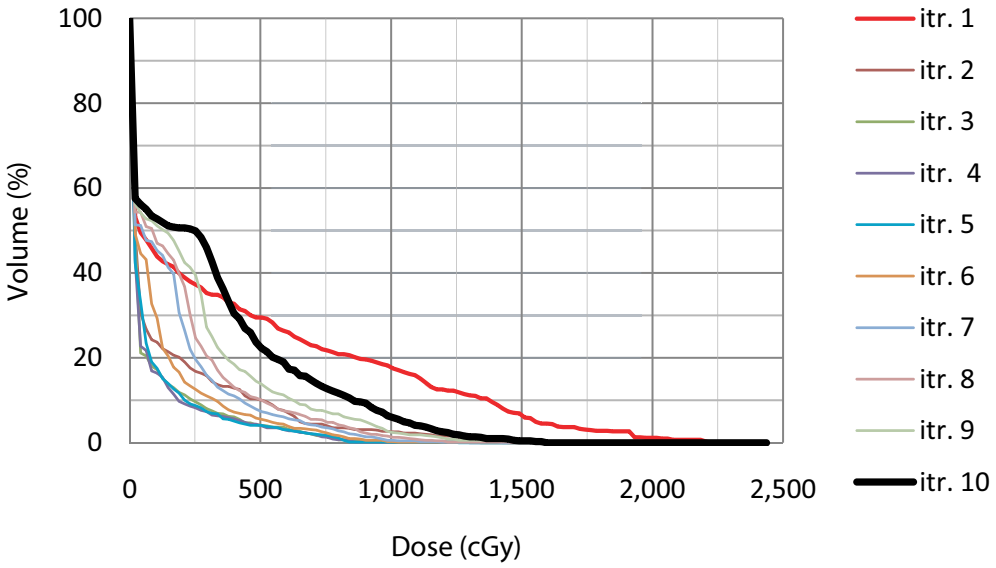

Figure 37: Change in Spinal Cord Dose-Volume-Histogram over the Iterations

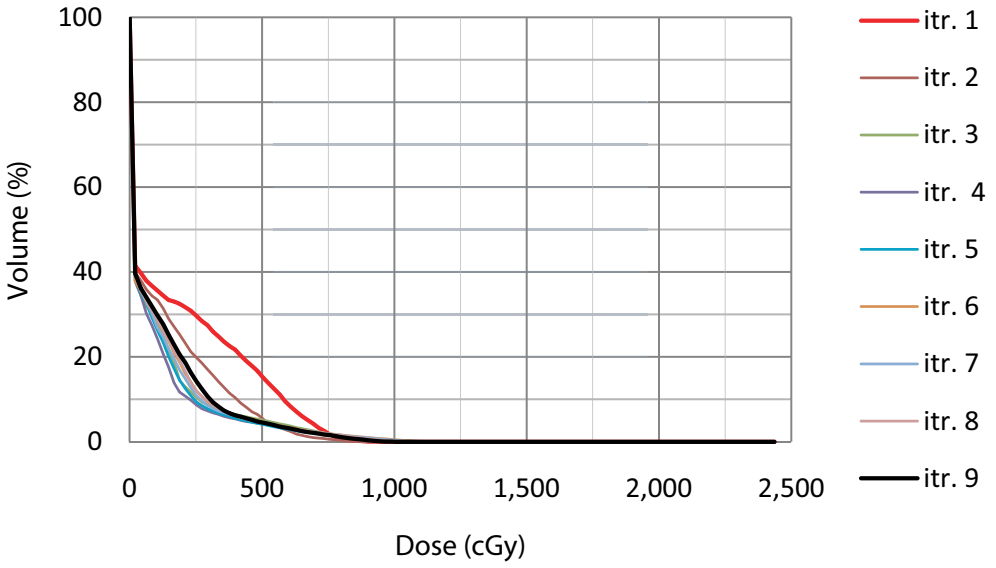

Figure 38: Change in Left Kidney Dose-Volume-Histogram over the Iterations

Table 21: Spine Case Performance Comparison for Valid Inequalities

	None	Eqn. (53)	Eqn. (54)	Eqn. (57)	Eqn. (53) Eqn. (54)	Eqn. (53) Eqn. (57)	Eqn. (54) Eqn. (57)	Eqn. (53) Eqn. (54) Eqn. (57)
Best Bound in 15 mins	262.47	261.27	288.15	289.69	288.33	292.63	292.78	295.48
Best Integer Solution in 15mins	333.61	356.36	N/A	298.28	N/A	300.44	**294.95**	298.9
Optimality Gap in 15 mins	21.32%	26.68%	N/A	2.88%	N/A	2.60%	0.74%	1.14%
Best Bound in 30 mins	262.47	266.05	288.15	290.23	288.33	293.12	293.07	295.81
Best Integer Solution in 30mins	333.61	356.36	315.84	296.35	N/A	299.43	**294.39**	297.89
Optimality Gap in 30 mins	21.32%	26.68%	8.77%	2.07%	N/A	2.11%	0.45%	0.69%
Best Bound in 45 mins	262.47	266.05	288.15	290.23	288.33	293.12	293.07	295.82
Best Integer Solution in 45mins	333.61	356.36	315.84	295.27	N/A	298.83	**294.17**	297.58
Optimality Gap in 45 mins	21.32%	26.68%	8.77%	1.71%	N/A	1.91%	0.37%	0.59%
Best Bound in 60 mins	262.47	266.05	289.69	290.23	288.33	293.12	293.07	295.82
Best Integer Solution in 60 mins	333.61	356.36	309.43	294.76	N/A	298.83	**294.03**	297.46
Optimality Gap in 60 mins	21.32%	26.68%	6.38%	1.54%	N/A	1.91%	0.32%	0.55%

Table 22: Prostate Case Performance Comparison for Valid Inequalities

	None	Eqn. (53)	Eqn. (54)	Eqn. (57)	Eqn. (53) Eqn. (54)	Eqn. (53) Eqn. (57)	Eqn. (54) Eqn. (57)	Eqn. (53) Eqn. (54) Eqn. (57)
Best Bound in 15 mins	119.84	119.96	122.76	122.99	122.93	123.19	123.25	123.41
Best Integer Solution in 15mins	N/A	N/A	126.38	125.07	N/A	125.16	**124.13**	124.16
Optimality Gap in 15 mins	N/A	N/A	2.86%	1.66%	N/A	1.52%	0.70%	0.61%
Best Bound in 30 mins	119.84	119.96	122.78	123.06	122.93	123.23	123.25	123.41
Best Integer Solution in 30mins	N/A	N/A	126.38	124.97	N/A	125.02	**123.98**	124.16
Optimality Gap in 30 mins	N/A	N/A	2.86%	1.52%	N/A	1.43%	0.58%	0.61%
Best Bound in 45 mins	119.84	119.96	122.86	123.06	122.93	123.23	123.25	123.4
Best Integer Solution in 45mins	N/A	N/A	125.04	124.74	126.75	124.57	**123.98**	124.09
Optimality Gap in 45 mins	N/A	N/A	1.74%	1.34%	3.02%	1.08%	0.58%	0.56%
Best Bound in 60 mins	119.84	119.96	122.86	123.06	122.97	123.23	123.25	123.4
Best Integer Solution in 60 mins	N/A	N/A	124.86	124.66	126.75	124.35	**123.96**	124.09
Optimality Gap in 60 mins	N/A	N/A	1.60%	1.28%	2.98%	0.90%	0.57%	0.56%

Table 23: Dose Prescriptions for the Spine Case

Constraint Type	Structure	Percentage	L (cGy)	U (cGy)
Full Volume	Spinal Cord	-	0	1575
	Kidney_LT	-	0	1000
	Kidney_RT	-	0	1000
	PTV	-	0	2415
Partial Volume	PTV	95	2100	-

Table 24: Dose Prescriptions for the Prostate Case

Constraint Type	Structure	Percentage	L (cGy)	U (cGy)
Full Volume	Rectum	-	0	8316
	Bladder	-	0	8316
	Femurs	-	0	5000
	PTV	-	0	8712
Partial Volume	PTV	95	7920	-

Table 25: The Values of Algorithm Parameters

t	N^t	E^t	CR^t	TR^t
1	15	5	40%	20%
2	12	4	35%	18%
3	9	3	30%	16%
4	6	2	25%	14%
5	3	1	20%	12%
6	3	1	15%	10%
7	3	1	10%	8%
8	3	1	5%	6%
9	3	1	0%	4%

Table 26: Solution Time in Seconds for the Spine Case

	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5	iter. 6	iter. 7	iter. 8	iter. 9
Limit	900	900	900	900	900	900	900	900	900
First Integer Solution	310	367	627	342	220	320	483	581	729
Reported Solution	310	553	897	342	775	742	483	581	729

Table 27: Solution Time in Seconds for the Prostate Case

	iter. 1	iter. 2	iter. 3	iter. 4	iter. 5	iter. 6	iter. 7	iter. 8	iter. 9
Limit	900	900	900	900	900	900	900	900	900
First Integer Solution	130	277	534	313	220	298	278	285	359
Reported Solution	130	351	534	313	853	298	278	285	359

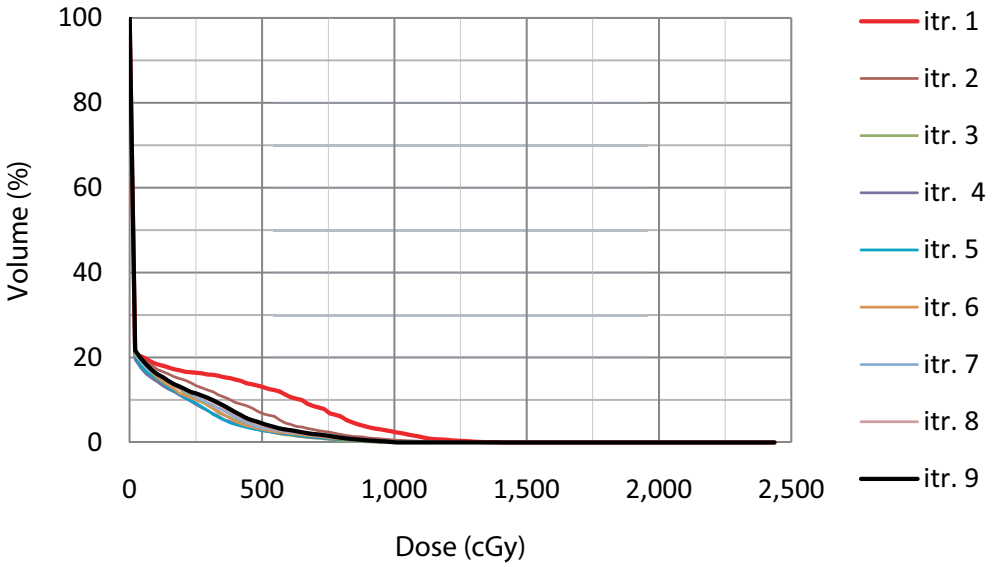

Figure 39: Change in Right Kidney Dose-Volume-Histogram over the Iterations

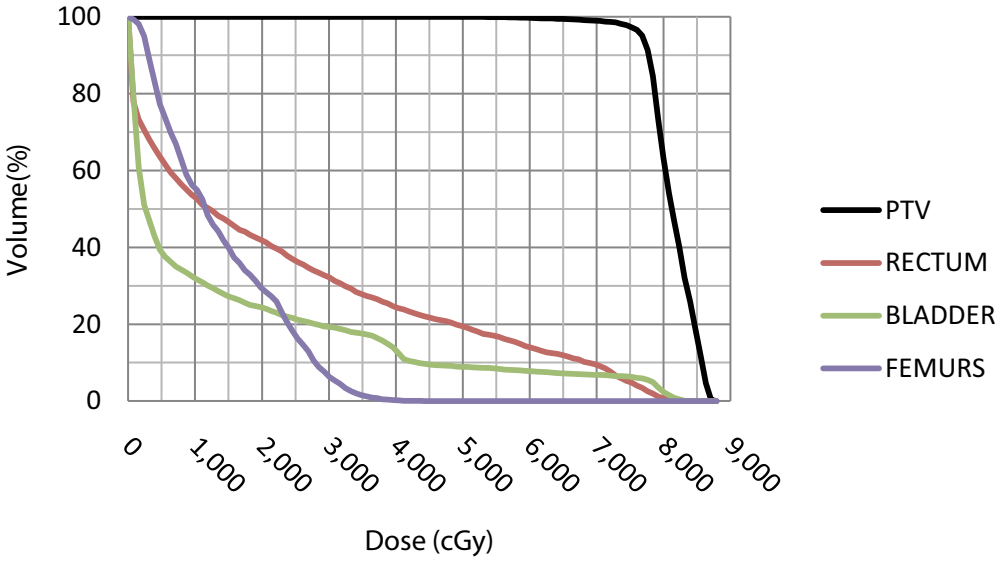

Figure 40: The Final Dose-Volume-Histogram for the Prostate Case

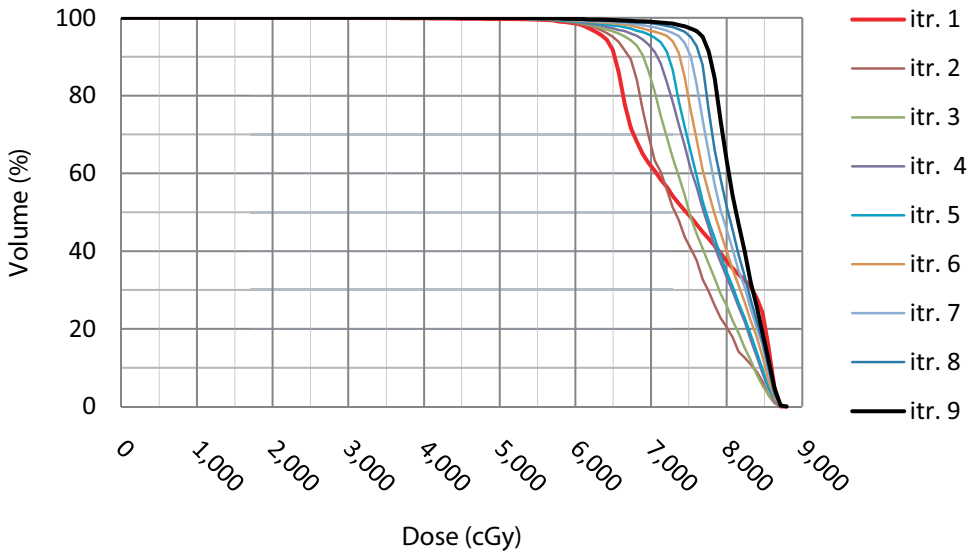

Figure 41: Change in PTV (Prostate Case) Dose-Volume-Histogram over the Iterations

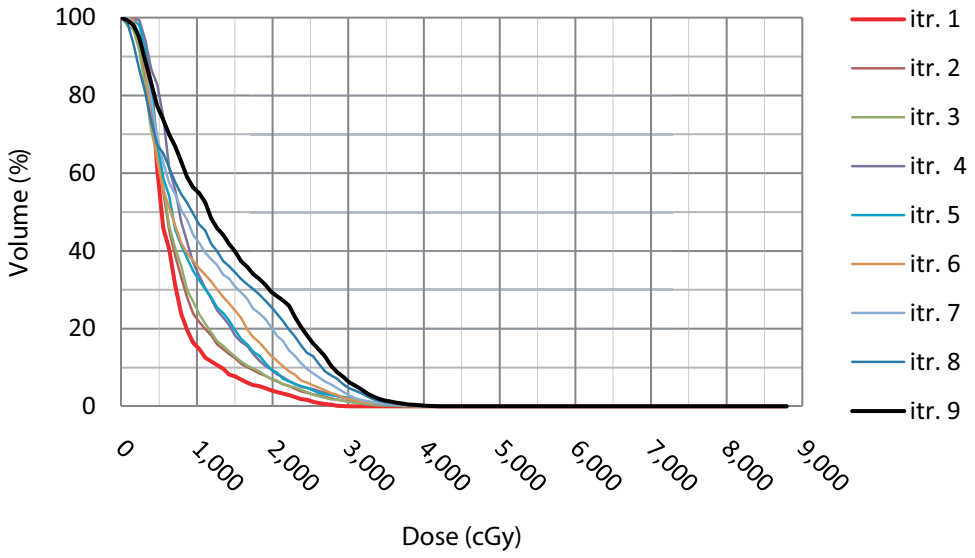

Figure 42: Change in Femurs Dose-Volume-Histogram over the Iterations

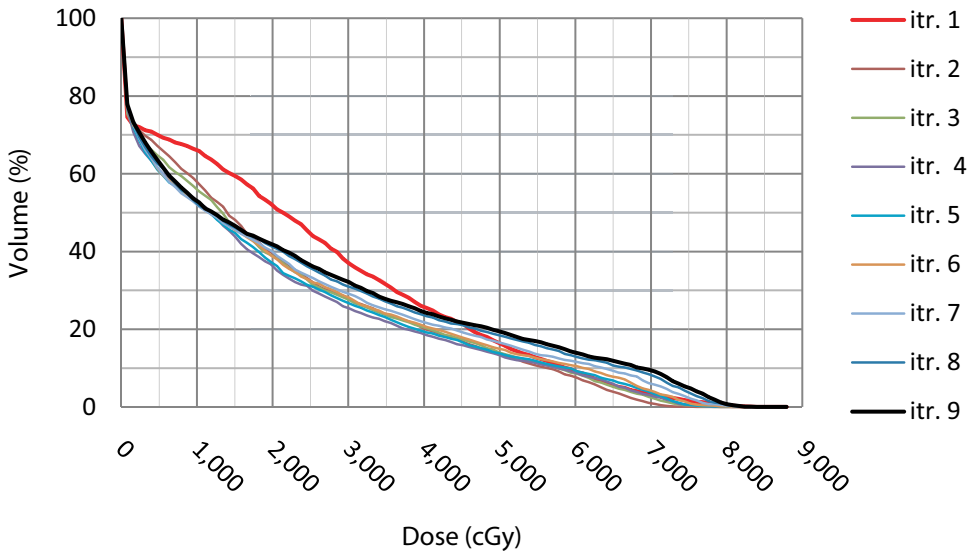

Figure 43: Change in Rectum Dose-Volume-Histogram over the Iterations

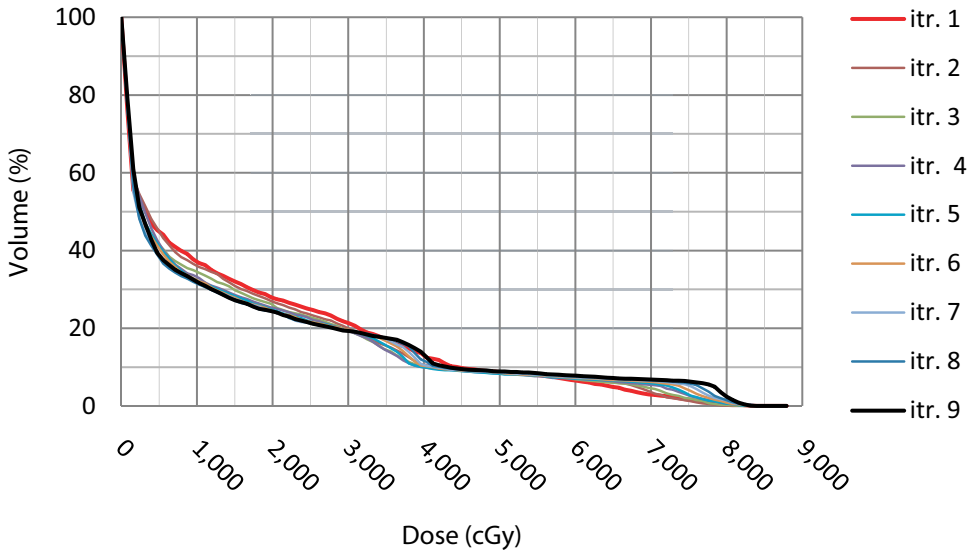

Figure 44: Change in Bladder Dose-Volume-Histogram over the Iterations

REFERENCES

[1] Ahnesjo, A., "Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media," *Medical Physics*, vol. 16, no. 4, pp. 577–592, 1989.

[2] Bahr, G. K., Kereiakes, J. G., Horwitz, H., Finney, R., Galvin, J., and Goode, K., "The method of linear programming applied to radiation treatment planning," *Radiology*, vol. 91, pp. 686–693, 1968.

[3] Bedford, J. L., Hansen, V. N., McNair, H. A., Aitken, A. H., Brock, J. E. C., Warrington, A. P., and Brada, M., "Treatment of lung cancer using volumetric modulated arc therapy and image guidance: A case study," *Acta Oncologica*, vol. 47, no. 7, pp. 1438–1443, 2008.

[4] Bedford, J. L. and Webb, S., "Direct-aperture optimization applied to selection of beam orientations in intensity-modulated radiation therapy," *Physics in Medicine and Biology*, vol. 52, no. 2, pp. 479–498, 2007.

[5] Bortfeld, T., "IMRT: a review and preview," *Physics in Medicine and Biology*, vol. 51, no. 13, pp. R363–R379, 2006.

[6] Boyer, A. L., Butler, E. B., DiPetrillo, T. A., Engler, M. J., Fraass, B., Grant, W., Ling, C. C., Low, D. A., Mackie, T. R., Mohan, R., Purdy, J. A., Roach, M., Rosenman, J. G., Verhey, L. J., Wong, J. W., Cumberlin, R. L., Stone, H., and Palta, J. R., "Intensity-modulated radiotherapy: Current status and issues of interest," *International Journal of Radiation Oncology Biology Physics*, vol. 51, no. 4, pp. 880–914, 2001.

[7] Cameron, C., "Sweeping-window arc therapy: an implementation of rotational imrt with automatic beam-weight calculation," *Physics in Medicine and Biology*, vol. 50, no. 18, pp. 4317–4336, 2005.

[8] Chui, C. S., Yorke, E., and Hong, L., "The effects of intra-fraction organ motion on the delivery of intensity-modulated field with a multileaf collimator," *Medical Physics*, vol. 30, no. 7, pp. 1736–1746, 2003.

[9] Clark, G. M., Popple, R. A., Young, P. E., and Fiveash, J. B., "Feasibility of single-isocenter volumetric modulated arc radiosurgery for treatment of multiple brain metastases," *International Journal of Radiation Oncology Biology Physics*, vol. 76, no. 1, pp. 296–302, 2010.

[10] Cozzi, L., Dinshaw, K. A., Shrivastava, S. K., Mahantshetty, U., Engineer, R., Deshpande, D. D., Jamema, S. V., Vanetti, E., Clivio, A., Nicolini, G., and Fogliata, A., "A treatment planning study comparing volumetric arc modulation with rapidarc and fixed field imrt for cervix uteri radiotherapy," *Radiotherapy and Oncology*, vol. 89, no. 2, pp. 180–191, 2008.

[11] D'Souza, W., Meyer, R., and Shi, L., "Selection of beam orientations in intensity-modulated radiation therapy using single-beam indices and integer programming," *Physics in Medicine and Biology*, vol. 49, pp. 3465–3481, AUG 7 2004.

[12] Earl, M. A., Shepard, D. M., Naqvi, S., Li, X. A., and Yu, C. X., "Inverse planning for intensity-modulated arc therapy using direct aperture optimization," *Physics in Medicine and Biology*, vol. 48, no. 8, pp. 1075–1089, 2003.

[13] Ehrgott, M., Holder, A., and Reese, J., "Beam selection in radiotherapy design," *Linear Algebra and Its Applications*, vol. 428, no. 5-6, pp. 1272–1312, 2008.

[14] Fogliata, A., Clivio, A., Nicolini, G., Vanetti, E., and Cozzi, L., "Intensity modulation with photons for benign intracranial tumours: A planning comparison of volumetric single arc, helical arc and fixed gantry techniques," *Radiotherapy and Oncology*, vol. 89, no. 3, pp. 254–262, 2008.

[15] Girinsky, T., Pichenot, C., Beaudre, A., Ghalibafian, M., and Lefkopoulos, D., "Is intensity-modulated radiotherapy better than conventional radiation treatment and three-dimensional conformal radiotherapy for mediastinal masses in patients with hodgkin's disease, and is there a role for beam orientation optimization and dose constraints assigned to virtual volumes?," *International Journal of Radiation Oncology Biology Physics*, vol. 64, no. 1, pp. 218–226, 2006.

[16] Holder, A., "Designing radiotherapy plans with elastic constraints and interior point methods," *Health Care and Management Science*, vol. 6, no. 1, pp. 5–16, 2003.

[17] Jiang, S. B., Pope, C., Al Jarrah, K. M., Kung, J. H., Bortfeld, T., and Chen, G. T. Y., "An experimental investigation on intra-fractional organ motion effects in lung imrt treatments," *Physics in Medicine and Biology*, vol. 48, no. 12, pp. 1773–1784, 2003.

[18] Keall, P. J., Joshi, S., Vedam, S. S., Siebers, J. V., Kini, V. R., and Mohan, R., "Four-dimensional radiotherapy planning for dmlc-based respiratory motion tracking," *Medical Physics*, vol. 32, no. 4, pp. 942–951, 2005.

[19] Kessler, M. L., McShan, D. L., Epelman, M. A., Vineberg, K. A., Eisbruch, A., Lawrence, T. S., and Fraass, B. A., "Costlets: A generalized approach to cost functions for automated optimization of imrt treatment plans," *Optimization and Engineering*, vol. 6, no. 4, pp. 421–448, 2005.

[20] Lee, E. K., Fox, T., and Crocker, I., "Integer programming applied to intensity-modulated radiation therapy treatment planning," *Annals of Operations Research*, vol. 119, no. 1-4, pp. 165–181, 2003.

[21] Lee, E. K., Fox, T., and Crocker, I., "Simultaneous beam geometry and intensity map optimization in intensity-modulated radiation therapy," *International Journal of Radiation Oncology Biology Physics*, vol. 64, no. 1, pp. 301–320, 2006.

[22] Lee, L., Ma, Y. Z., Ye, Y. Y., and Xing, L., "Conceptual formulation on four-dimensional inverse planning for intensity modulated radiation therapy," *Physics in Medicine and Biology*, vol. 54, no. 13, pp. N255–N266, 2009.

[23] Li, X. A., Stepaniak, C., and Gore, E., "Technical and dosimetric aspects of respiratory gating using a pressure-sensor motion monitoring system," *Medical Physics*, vol. 33, no. 1, pp. 145–154, 2006.

[24] Lim, G. J., Choi, J., and Mohan, R., "Iterative solution methods for beam angle and fluence map optimization in intensity modulated radiation therapy planning," *OR Spectrum*, vol. 30, no. 2, pp. 289–309, 2008.

[25] Liu, H. H., Balter, P., Tutt, T., Choi, B., Zhang, J., Wang, C., Chi, M., Luo, D. S., Pan, T. S., Hunjan, S., Starkschall, G., Rosen, I., Prado, K., Liao, Z. X., Chang, J., Komaki, R., Cox, J. D., Mohan, R., and Dong, L., "Assessing respiration-induced tumor motion and internal target volume using four-dimensional computed tomography for radiotherapy of lung cancer," *International Journal of Radiation Oncology Biology Physics*, vol. 68, no. 2, pp. 531–540, 2007.

[26] Ma, Y. Z., Popple, R., Suh, T. S., and Xing, L., "Beams-eye-view dosimetrics-guided inverse planning for aperture-modulated arc therapy," *International Journal of Radiation Oncology Biology Physics*, vol. 75, no. 5, pp. 1587–1595, 2009.

[27] Mackie, T. R., Scrimger, J. W., and Battista, J. J., "A convolution method of calculating dose for 15-mv x-rays," *Medical Physics*, vol. 12, no. 2, pp. 188–196, 1985.

[28] Mohan, R., Chui, C., and Lidofsky, L., "Differential pencil beam dose computation model for photons," *Medical Physics*, vol. 13, no. 1, pp. 64–73, 1986.

[29] Otto, K., "Volumetric modulated arc therapy: Imrt in a single gantry arc," *Medical Physics*, vol. 35, no. 1, pp. 310–317, 2008.

[30] Palma, D., Vollans, E., James, K., Nakano, S., Moiseenko, V., Shaffer, R., McKenzie, M., Morris, J., and Otto, K., "Volumetric modulated arc therapy for delivery of prostate radiotherapy: comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 72, no. 4, pp. 996–1001, 2008.

[31] Pan, T., Lee, T. Y., Rietzel, E., and Chen, G. T. Y., "4d-ct imaging of a volume influenced by respiratory motion on multi-slice ct," *Medical Physics*, vol. 31, no. 2, pp. 333–340, 2004.

[32] Price, R. A., Murphy, S., McNeeley, S. W., Ma, C. M. C., Horwitz, E., Movsas, B., Raben, A., and Pollack, A., "A method for increased dose conformity and segment reduction for smlc delivered imrt treatment of the prostate," *International Journal of Radiation Oncology Biology Physics*, vol. 57, no. 3, pp. 843–852, 2003.

[33] Pugachev, A. and Lei, X., "Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy," *International Journal of Radiation Oncology Biology Physics*, vol. 51, no. 5, pp. 1361–1370, 2001.

[34] Romeijn, H. E., Ahuja, R. K., Dempsey, J. F., and Kumar, A., "A new linear programming approach to radiation therapy treatment planning problems," *Operations Research*, vol. 54, no. 2, pp. 201–216, 2006.

[35] Rosu, M., Chetty, I. J., Balter, J. M., Kessler, M. L., McShan, D. L., and Ten Haken, R. K., "Dose reconstruction in deforming lung anatomy: Dose grid size effects and clinical implications," *Medical Physics*, vol. 32, no. 8, pp. 2487–2495, 2005.

[36] Rosu, M., Chetty, I. J., Tatro, D. S., and Ten Haken, R. K., "The impact of breathing motion versus heterogeneity effects in lung cancer treatment planning," *Medical Physics*, vol. 34, no. 4, pp. 1462–1473, 2007.

[37] Schreibmann, E., Thorndyke, B., Li, T., Wang, J., and Xing, L., "Four-dimensional image registration for image-guided radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 71, no. 2, pp. 578–586, 2008.

[38] Shaffer, R., Morris, W. J., Moiseenko, V., Welsh, M., Crumley, C., Nakano, S., Schmuland, M., Pickles, T., and Otto, K., "Volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for simultaneous maximal intraprostatic boost: a planning comparison study," *Clinical Oncology*, vol. 21, no. 5, pp. 401–407, 2009.

[39] Shepard, D. M., Cao, D., Afghan, M. K. N., and Earl, M. A., "An arc-sequencing algorithm for intensity modulated arc therapy," *Medical Physics*, vol. 34, no. 2, pp. 464–470, 2007.

[40] Shepard, D. M., Ferris, M. C., Olivera, G. H., and Mackie, T. R., "Optimizing the delivery of radiation therapy to cancer patients," *Siam Review*, vol. 41, no. 4, pp. 721–744, 1999.

[41] Shimizu, S., Shirato, H., Ogura, S., Akita-Dosaka, H., Kitamura, K., Nishioka, T., Kagei, K., Nishimura, M., and Miyasaka, K., "Detection of lung tumor movement in real-time tumor-tracking radiotherapy," *International Journal of Radiation Oncology Biology Physics*, vol. 51, no. 2, pp. 304–310, 2001.

[42] Stein, J., Mohan, R., Wang, X., Bortfeld, T., Wu, Q., Preiser, K., Ling, C., and Schlegel, W., "Number and orientations of beams in intensity-modulated radiation treatments," *Medical Physics*, vol. 24, pp. 149–160, FEB 1997.

[43] Suh, Y. L., Sawant, A., Venkat, R., and Keall, P. J., "Four-dimensional imrt treatment planning using a dmlc motion-tracking algorithm," *Physics in Medicine and Biology*, vol. 54, no. 12, pp. 3821–3835, 2009.

[44] Ulrich, S., Nill, S., and Oelfke, U., "Development of an optimization concept for arc-modulated cone beam therapy," *Physics in Medicine and Biology*, vol. 52, no. 14, pp. 4099–4119, 2007.

[45] Vedam, S. S., Keall, P. J., Kini, V. R., Mostafavi, H., Shukla, H. P., and Mohan, R., "Acquiring a four-dimensional computed tomography dataset using an external respiratory signal," *Physics in Medicine and Biology*, vol. 48, no. 1, pp. 45–62, 2003.

[46] Veldeman, L., Madani, I., Hulstaert, F., De Meerleer, G., Mareel, M., and De Neve, W., "Evidence behind use of intensity-modulated radiotherapy: a systematic review of comparative clinical studies," *Lancet Oncology*, vol. 9, no. 4, pp. 367–375, 2008.

[47] Wagman, R., Yorke, E., Ford, E., Giraud, P., Mageras, G., Minsky, B., and Rosenzweig, K., "Respiratory gating for liver tumors: Use in dose escalation," *International Journal of Radiation Oncology Biology Physics*, vol. 55, no. 3, pp. 659–668, 2003.

[48] Wang, C., Luan, S., Tang, G., Chen, D. Z., Earl, M. A., and Yu, C. X., "Arc-modulated radiation therapy (amrt): a single-arc form of intensity-modulated arc therapy," *Physics in Medicine and Biology*, vol. 53, no. 22, pp. 6291–6303, 2008.

[49] Webb, S., "The physical basis of imrt and inverse planning," *British Journal of Radiology*, vol. 76, no. 910, pp. 678–689, 2003.

[50] Xi, M., Liu, M. Z., Deng, X. W., Zhang, L., Huang, X. Y., Liu, H., Li, Q. Q., Hu, Y. H., Cai, L., and Cui, N. J., "Defining internal target volume (itv) for hepatocellular carcinoma using four-dimensional ct," *Radiotherapy and Oncology*, vol. 84, no. 3, pp. 272–278, 2007.

[51] Yang, R. J., Dai, J. R., Yang, Y., and Hu, Y. M., "Beam orientation optimization for intensity-modulated radiation therapy using mixed integer programming," *Physics in Medicine and Biology*, vol. 51, no. 15, pp. 3653–3666, 2006.

[52] Yang, Y. and Xing, L., "Clinical knowledge-based inverse treatment planning," *Physics in Medicine and Biology*, vol. 49, no. 22, pp. 5101–5117, 2004.

[53] Yu, C. X., "Intensity-modulated arc therapy with dynamic multileaf collimation: an alternative to tomotherapy," *Physics in Medicine and Biology*, vol. 40, pp. 1435–1449, 1995.

[54] Yu, C. X., Jaffray, D. A., and Wong, J. W., "The effects of intra-fraction organ motion on the delivery of dynamic intensity modulation," *Physics in Medicine and Biology*, vol. 43, no. 1, pp. 91–104, 1998.